LOW-RANK MATRIX RECOVERY VIA ITERATIVELY REWEIGHTED LEAST SQUARES MINIMIZATION

MASSIMO FORNASIER,* HOLGER RAUHUT,† AND RACHEL WARD‡

Abstract. We present and analyze an efficient implementation of an iteratively reweighted least squares algorithm for recovering a matrix from a small number of linear measurements. The algorithm is designed for the simultaneous promotion of both a minimal *nuclear norm* and an approximatively *low-rank* solution. Under the assumption that the linear measurements fulfill a suitable generalization of the *Null Space Property* known in the context of compressed sensing, the algorithm is guaranteed to recover iteratively any matrix with an error of the order of the best k-rank approximation. In certain relevant cases, for instance for the *matrix completion* problem, our version of this algorithm can take advantage of the *Woodbury matrix identity*, which allows to expedite the solution of the least squares problems required at each iteration. We present numerical experiments which confirm the robustness of the algorithm for the solution of matrix completion problems, and demonstrate its competitiveness with respect to other techniques proposed recently in the literature.

AMS subject classification: 65J22, 65K10, 52A41, 49M30.

Key Words: low-rank matrix recovery, iteratively reweighted least squares, matrix completion.

1. Introduction. Affine rank minimization refers to the problem of finding a matrix of minimal rank consistent with a given underdetermined linear system of equations. This problem arises in many areas of science and technology, including system identification [27], collaborative filtering, quantum state tomography [18, 17], signal processing, and image processing. An important special case is the matrix completion problem [6, 7, 33], where one would like to fill in missing entries of a large data matrix which is assumed to have low-rank.

Unfortunately, the affine rank minimization problem is NP-hard in general [34, 28]; therefore, it is desirable to have tractable alternatives. In [11], Fazel studied *nuclear norm* minimization for this purpose, which was known to be a good proxy for rank minimization. The nuclear norm of a matrix is the ℓ_1-norm of its singular values, and is the smallest convex envelope of the rank function [11]. Reformulated as a semidefinite program, nuclear norm minimization can be solved with efficient methods [3].

1.1. Contribution of this paper. Unfortunately, standard semidefinite programming tools work efficiently for solving nuclear norm minimization problems only for matrices up to size approximately 100×100. Therefore, it is crucial to develop fast algorithms that are specialized to nuclear norm minimization (or other heuristics for rank minimization). So far, several approaches have been suggested [23, 4, 15, 20, 21]. Some aim at general nuclear norm minimization problems and others are specialized to matrix completion problems. Inspired by the iteratively reweighted least-squares algorithm analyzed in [10] for sparse vector recovery, we develop a variant of this algorithm for nuclear norm minimization / low-rank matrix recovery. Each step of the proposed algorithm requires the computation of a partial singular value decomposition and the solution of a (usually small) least squares problem, and both of these tasks can be performed efficiently. The analysis of our algorithm is based on a suitable

---

*JOHANN RADON INSTITUTE FOR COMPUTATIONAL AND APPLIED MATHEMATICS, AUSTRIAN ACADEMY OF SCIENCES, ALTENBERGERSTRASSE 69, A-4040 LINZ, AUSTRIA, EMAIL: MASSIMO.FORNASIER@OEAW.AC.AT.

†HAUSDORFF CENTER FOR MATHEMATICS & INSTITUTE FOR NUMERICAL SIMULATION, UNIVERSITY OF BONN, ENDENICHER ALLEE 60, D-53115 BONN, GERMANY, EMAIL: RAUHUT@HCM.UNI-BONN.DE.

‡MATHEMATICS DEPARTMENT, COURANT INSTITUTE OF MATHEMATICAL SCIENCES, NEW YORK UNIVERSITY, 251 MERCER STREET, NEW YORK, N.Y. 10012-1185, U.S.A., EMAIL: RWARD@CIMS.NYU.EDU.

matrix extension of the *null space property*, well-known in the approximation theory literature in connection with ℓ_1-minimization, see [31, 30] and reference therein, and recently popularized in the context of *compressive sensing* [12]. We show that the algorithm essentially has the same recovery guarantees as nuclear norm minimization. Numerical experiments also show that our algorithm is competitive with other state-of-the-art algorithms in the matrix completion setup [20, 21, 15]. Unfortunately, the null space property fails in the matrix completion setup, and the theoretical analysis of the algorithm seems to be much more involved. Such a theoretical analysis will be postponed to later investigations.

1.2. Low-rank matrix recovery and applications. In the following, we will deal with real or complex matrices indifferently and we denote the space of $n \times p$ matrices by $M_{n \times p}$. Given a linear map $\mathscr{S} : M_{n \times p} \to \mathbb{C}^m$, with $m \ll pn$, and a vector $\mathscr{M} \in \mathbb{C}^m$, we consider the affine rank minimization problem

$$\min_{X \in M_{n \times p}} \operatorname{rank}(X) \quad \text{subject to } \mathscr{S}(X) = \mathscr{M}. \tag{1.1}$$

An important special case of low-rank matrix recovery is matrix completion, where $\mathscr{S}$ samples entries,

$$\mathscr{S}(X)_\ell = x_{ij}, \tag{1.2}$$

for some i, j depending on ℓ. Such low-rank matrix recovery problems often arise; examples include the *Netflix problem*[1] or the recovery of positions from partial distance information [5]. We refer to [6, 7] for further details. As an example of low-rank matrix recovery from more general linear measurements, we can turn to quantum state tomography [18], where one tries to recover a quantum state, that is, a square matrix $X \in M_{n \times n}$, that is positive semidefinite and has trace 1. A pure state has rank 1, and a mixed state is of low-rank, or approximately low-rank. Then one has given a collection of matrices $A_k \in M_{n \times n}$, $k = 1, \ldots, n^2$, (so called Pauli-Matrices) and takes partial "quantum observations" $\mathscr{M}_j = \operatorname{Tr}(A_{k_j}^* X)$, $j = 1, \ldots, r$ with $r \leq n^2$, and the task is to recover a low-rank state X. We refer to [18, 17] for details.

1.3. Theoretical results. As already mentioned the rank minimization problem (1.1) is NP-hard in general, and therefore we consider its convex relaxation

$$\min_{X \in M_{n \times p}} \|X\|_* \quad \text{subject to } \mathscr{S}(X) = \mathscr{M}, \tag{1.3}$$

where $\|X\|_* = \sum_{i=1}^n \sigma_i(X)$ denotes the nuclear norm (or Schatten-1 norm, or trace norm), where $\sigma_i(X)$ are the singular values of X. There are two known regimes where nuclear norm minimization can be guaranteed to return minimal-rank solutions:

1. RIP measurement maps. The (rank) restricted isometry property (RIP) is analogous to the by-now classical restricted isometry property (RIP) from compressed sensing [5, 34].

DEFINITION 1.1 (Restricted Isometry Property [34]). *Let* $\mathscr{S} : M_{n \times p} \to \mathbb{C}^m$ *be a linear map. For every integer* k, *with* $1 \leq k \leq n$, *define the* k-*restricted isometry constant* $\delta_k = \delta_k(\mathscr{S}) > 0$ *to be the smallest number such that*

$$(1 - \delta_k)\|X\|_F^2 \leq \|\mathscr{S}(X)\|_{\ell_2^m}^2 \leq (1 + \delta_k)\|X\|_F^2$$

holds for all k-*rank matrices* X.

It is shown in [34] that nuclear norm miminization (1.3) recovers all matrices X of rank at most k from the measurements $\mathscr{M} = \mathscr{S}(X)$ provided δ_{5k} is small enough. Below we improve this to $\delta_{4k} \leq \sqrt{2} - 1$, and we refer to [5, 29] for related results.

---

[1]Netflix Prize sought to substantially improve the accuracy of predictions about how much someone is going to enjoy a movie based on their movie preferences `http://www.netflixprize.com/`.

The restricted isometry property is known to hold for Gaussian (or more generally sub-gaussian) measurement maps $\mathscr{S}$ in [34, 5]. These are maps of the form

$$\mathscr{S}(X)_\ell = \sum_{k,j} a_{\ell,k,j} X_{k,j}, \quad \ell = 1, \ldots, m, \tag{1.4}$$

where the $a_{\ell,k,j}$ are independent normal distributed random variables with mean zero and variance $1/m$. Such a map satisfies $\delta_k \leq \delta \in (0,1)$ with high probability provided

$$m \geq C_\delta \max\{p, n\} k, \tag{1.5}$$

see Theorem 2.3 in [5]. Since the degrees of freedom of a rank k matrix $X \in M_{n \times p}$ are $k(n + p - k)$, the above bound matches this number up to possibly a constant. Therefore, the bound (1.5) is optimal.

Using recent results in [1, 22], it follows that the restricted isometry property also holds for certain *structured* random matrices if slightly more measurements are allowed; in particular, for maps $\mathscr{S}$ of the form (1.4) where the $a_{\ell,k,j}$ – unraveled for each ℓ into vectors of length pn – correspond to m rows randomly selected from the $pn \times pn$ discrete Fourier matrix (or 2D Fourier transform matrix) with randomized column signs, $\delta_k \leq \delta \in (0,1)$ with high probability as long as

$$m \geq C_\delta \max\{p, n\} k \log^4(pn). \tag{1.6}$$

This follows from the recent findings in [1, 22] that such random partial Fourier measurements satisfy a concentration inequality of the form, for all $0 < \varepsilon < 1$,

$$\mathbb{P}\Big(\big| \|\mathscr{S}(X)\|^2 - \|X\|_F^2 \big| \geq \varepsilon \|X\|_F^2 \Big) \leq 2 \exp\big(-\frac{m}{2} C_\varepsilon \log^{-4}(pn)\big). \tag{1.7}$$

Subgaussian measurement maps also satisfy (1.7), and for these maps, the factor of $\log^{-4}(pn)$ in the exponent can be removed. The proof of RIP for subgaussian random ensembles appearing in Theorem 2.3 in [5] (see also Theorem 4.2 in [34]) relies only on such concentration for subgaussian measurement maps. Therefore, obvious modifications to that proof to accommodate the additional logarithmic factors in the exponent (1.7) give the stated RIP results for random partial Fourier measurements.

2. Matrix completion. In the matrix completion setup (1.2) where measurements are point-wise observations of entries of the matrix, there are obvious rank one matrices in the kernel of the operator $\mathscr{S}$; therefore, the RIP fails completely in this setting, and 'localized' low-rank matrices in the null space of $\mathscr{S}$ cannot be recovered by any method whatsoever. However, if certain conditions on the left and right singular vectors of the underlying low-rank matrix are imposed, essentially requiring that such vectors are uncorrelated with the canonical basis, then it was shown in [6, 7, 33] that such incoherent matrices of rank at most k can be recovered from m randomly chosen entries with high probability provided

$$m \geq Ck \max\{n, p\} \log^2(\max\{n, p\}).$$

Up to perhaps the exponent 2 at the log-term, this is optimal. We refer to [6, 7, 33] for detailed statements. In [18, 17] extensions to quantum state tomography can be found.

2. Notation and Preliminaries.

2.1. Notation. The entries of a matrix $X \in M_{n \times p}$ are denoted by lower case letters and the corresponding indices, i.e., $X_{ij} = x_{ij}$. We denote the adjoint matrix $X^* \in M_{p \times n}$ as usual.

In the following and without loss of generality, we assume that $n \leq p$. For a generic matrix $X \in M_{n \times p}$ we write its *singular value decomposition*

$$X = U\Sigma V^*$$

for unitary matrices $U \in M_{n \times n}, V \in M_{p \times p}$ and a diagonal matrix $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_n) \in M_{n \times p}$, where $\sigma_1 \geq \sigma_2 \geq \ldots \sigma_n \geq 0$ are the *singular values*. In the following, we denote $\sigma(X)$ the vector of the singular values of the matrix X. For a specific matrix X we indicate the singular value decomposition as $X = U_X \Sigma_X V_X^*$. For self-adjoint matrices $X = X^*$, we have $V = U$. We write $X \succ 0$ if X is self-adjoint and positive-definite. In this case we can define, for $\alpha \in \mathbb{R}$, the α−th power of the matrix X by $X^\alpha = U\Sigma^\alpha U^*$, where $\Sigma^\alpha = \text{diag}(\sigma_1^\alpha, \ldots, \sigma_n^\alpha)$. If X is positive semi-definite, we write $X \succeq 0$.

The rank of $X \in M_{n \times p}$ denoted by $\text{rank}(X)$ equals the number of nonzero singular values of X. We will say that X is a *k-rank matrix* if $\text{rank}(X) \leq k$. The trace of a square matrix $X \in M_{n \times n}$ is the sum of its diagonal entries, i.e., $\text{Tr}(X) = \sum_{i=1}^n x_{ii}$. The trace satisfies $\text{Tr}(X) = \overline{\text{Tr}(X^*)}$ and is cyclic, that is, $\text{Tr}(XY) = \text{Tr}(YX)$ for all matrices X, Y with matching dimensions. If λ_i are the eigenvalues of X then $\text{Tr}(X) = \sum_i \lambda_i$. Denote by E_{ij} the matrix whose (i, j)-entry is 1 and all other entries take the value 0. Then $\text{Tr}(E_{ij}X) = x_{ji}$.

The space of complex matrices $M_{n \times p}$ forms a Hilbert space when endowed with the natural scalar product $\langle X, Y \rangle = \text{Tr}(XY^*)$, which induces the *Frobenius norm* $\|X\|_F = \langle X, X \rangle^{1/2}$. The following identities hold: $\|X\|_F = \left(\sum_{i=1}^n \sum_{j=1}^p |x_{ij}|^2 \right)^{1/2} = \left(\sum_{i=1}^n \sigma_i^2 \right)^{1/2}$. More generally, we consider the Schatten q-norms $\|X\|_{S_q} := \|\sigma(X)\|_{\ell_q^n}$ [2, 19], which are based on the ℓ_q^n-vector norms $\|x\|_{\ell_q^n} := \left(\sum_{i=1}^n |x_i|^q \right)^{1/q}$, $1 \leq q < \infty$, and $\|x\|_{\ell_\infty^n} := \max_{i=1}^n |x_i|$. Of particular importance is the Schatten 1-norm, or nuclear norm, which we also denote by $\|X\|_* = \|X\|_{S_1} = \sum_{i=1}^n \sigma_i(X)$. The operator norm $\|X\|_{op} = \|X\|_{S_\infty}$ will be needed as well.

If XX^* is invertible, then the *absolute value* of X is defined by

$$|X| = (XX^*)^{-1/2}XX^* = (XX^*)^{1/2}; \tag{2.1}$$

in this case, the Schatten q-norm satisfies

$$\|X\|_{S_q} := (\text{Tr} |X|^q)^{1/q} \tag{2.2}$$

for all $1 \leq q < \infty$.

We consider linear operators on matrices of the type $\mathscr{S} : M_{n \times p} \to \mathbb{C}^m$. We denote the action of $\mathscr{S}$ on the matrix X by $\mathscr{S}(X)$ in order to distinguish it from $\mathscr{S}X$ which may create confusion and be interpreted as a matrix-matrix multiplication in certain passages below. Rather, as the vector space $M_{n \times p}$ is isomorphic to the vector space $\mathbb{C}^{n \times p}$, linear operators of the form $\mathscr{S} : M_{n \times p} \to \mathbb{C}^m$ may be interpreted as elements of $M_{m \times np}$. We denote by $\mathscr{S}^*$ the adjoint operator of $\mathscr{S}$, such that $\langle \mathscr{S}(X), \lambda \rangle_{\mathbb{C}^m} = \langle X, \mathscr{S}^*(\lambda) \rangle$ for all $X \in M_{n \times p}, \lambda \in \mathbb{C}^m$, where the former scalar product is the Euclidean one on $\mathbb{C}^m$ and the latter scalar product is the one inducing the Frobenius norm. We denote by $I := I_n \in M_{n \times n}$ the identity matrix.

Finally, let us define the *k-spectral truncation* of X by

$$X_{[k]} = U\Sigma_{[k]}V^*,$$

where $\Sigma_{[k]} = \text{diag}(\sigma_1, \ldots, \sigma_k, 0, \ldots, 0)$. Thus, this operation acts by setting to 0 all the singular values with indexes from $k + 1$ to n. We also introduce the ε-stabilization of X by

$$X_\varepsilon = U\Sigma_\varepsilon V^*, \tag{2.3}$$

where $\Sigma_\varepsilon = \text{diag}(\max\{\sigma_j, \varepsilon\})$. That is, under this operation, all singular values below a certain minimal threshold are increased to the threshold.

For a self-adjoint matrix $X = U\Sigma U^* \in M_{n \times n}$, note that

$$X_\varepsilon = U\widetilde{\Sigma}U^* + \varepsilon I, \tag{2.4}$$

where $\widetilde{\Sigma} = \text{diag}\left(\max\{\sigma_j - \varepsilon, 0\}\right)$. In particular, if a significant number of the singular values of X fall below the threshold ε, then X_ε is decomposed into the sum of a low-rank matrix and a scaled identity matrix.

2.2. Motivation and formulation of the algorithm. Let us fix a sampling operator $\mathscr{S} : M_{n \times p} \to \mathbb{C}^m$ and a measured data vector $\mathscr{M} \in \mathbb{C}^m$. We are interested in the rank minimization problem

$$\arg \min_{\mathscr{S}(X) = \mathscr{M}} \text{rank}(X). \tag{2.5}$$

As mentioned in the introduction, this problem can be efficiently solved under suitable assumptions on $\mathscr{S}$ by considering its convex relaxation

$$\arg \min_{\mathscr{S}(X) = \mathscr{M}} \|X\|_*, \tag{2.6}$$

where rank minimization is substituted by nuclear norm minimization.

We propose an algorithm for solving (2.6) reminiscent of iteratively reweighted least squares algorithm for linearly-constrained ℓ_1-norm minimization, and is based on the following motivation: if all of the singular values of $X \in M_{n \times p}$ are nonzero, then, according to (2.1)-(2.2), we may re-write its nuclear norm as

$$\|X\|_* = \|X\|_{S^1} = \text{Tr}\left[(XX^*)^{-1/2}(XX^*)\right] = \|W^{1/2}X\|_F^2, \tag{2.7}$$

where $W = (XX^*)^{-1/2}$. This suggests the following approach for solving the nuclear norm minimization problem (2.6): let us initialize a weight matrix $W^0 \in M_{n \times n}$, and then recursively define, for $\ell = 0, 1, \ldots,$

$$X^{\ell+1} = \arg \min_{\mathscr{S}(X) = \mathscr{M}} \|(W^\ell)^{1/2}X\|_F^2, \quad \text{and} \quad W^{\ell+1} = \left(X^{\ell+1}(X^{\ell+1})^*\right)^{-1/2}. \tag{2.8}$$

Observe that the minimization $\bar{X} = \arg \min_{\mathscr{S}(X) = \mathscr{M}} \|W^{1/2}X\|_F^2$ can be reformulated as a weighted least squares problem with m linear constraints on the np variables x_{ij}; each of the updates for $X^{\ell+1}$ and $W^{\ell+1}$ can then be performed explicitly. However, once any of the singular values of $X^{\ell+1}$ become small – a desirable situation as we seek to recover low-rank matrices – the computation of $W^{\ell+1}$ becomes ill-conditioned.

To stabilize the algorithm, we fix a parameter $\varepsilon > 0$, and increase the small singular values of X^ℓ according to $\sigma_i(X^\ell) \longrightarrow \max\{\sigma_i(X^\ell), \varepsilon\}$ after updating X^ℓ but before computing W^ℓ. That is, we replace X^ℓ by its ε-stabilization $X_\varepsilon^\ell = U\Sigma_\varepsilon^\ell V^*$, as defined in (2.3). The idea is that X_ε^ℓ is well-conditioned and, additionally, if X^ℓ is well-approximated by a k-rank matrix, then X_ε^ℓ is well-approximated by the same k-rank matrix. Keeping $\varepsilon > 0$ fixed throughout the iterations, we would no longer expect the algorithm to converge to the nuclear norm solution (2.6). Nevertheless, if we allow for $\varepsilon = \varepsilon_\ell \to 0$ in such a way that $\varepsilon_\ell < \sigma_k(X^\ell)$ is maintained at each iteration, then one may hope for both stability and convergence towards a k-rank solution. Collecting these ideas together, we arrive at the following *iterative reweighted least squares algorithm* for low-rank matrix recovery (IRLS-M):

> **IRLS-M algorithm for low-rank matrix recovery:** Initialize $W^0 = I \in M_{n \times n}$. Set $\varepsilon_0 := 1$, $K \geq k$, and $\gamma > 0$. Then recursively define, for $\ell = 1, 2, \ldots,$
>
> $$X^\ell := \arg \min_{\mathscr{S}(X) = \mathscr{M}} \|(W^{\ell-1})^{1/2} X\|_F^2, \tag{2.9}$$
>
> $$\varepsilon_\ell := \min \left\{ \varepsilon_{\ell-1}, \gamma \sigma_{K+1}(X^\ell) \right\}. \tag{2.10}$$
>
> Compute $U^\ell \in M_{n \times n}$ and $\Sigma^\ell = \operatorname{diag}(\sigma_1^\ell, \ldots, \sigma_n^\ell) \in M_{n \times n}$ for which
>
> $$X^\ell (X^\ell)^* = U^\ell (\Sigma^\ell)^2 (U^\ell)^*. \tag{2.11}$$
>
> Set $\qquad W^\ell = U^\ell (\Sigma_{\varepsilon_\ell}^\ell)^{-1} (U^\ell)^*, \tag{2.12}$
>
> where Σ_ε denotes the ε-stabilization of the matrix Σ, see (2.3). The algorithm stops if $\varepsilon_\ell = 0$; in this case, define $X^j := X^\ell$ for $j > \ell$. In general, the algorithm generates an infinite sequence $(X^\ell)_{\ell \in \mathbb{N}}$ of matrices.

Reduction to iteratively reweighted least squares for vector recovery. Suppose now that $n = p$ and the linear constraints $\mathscr{S}(X) = \mathscr{M}$ are equivalent to the $m \times n$ linear system $Sx = \mathscr{M}$ exclusively acting on the vector $x = \operatorname{diag} X$, and $x_{ij} = 0$ is enforced for $i \neq j$. Then the nuclear norm minimization problem (2.6) reduces to that of finding the vector $x \in \mathbb{C}^n$ of minimal ℓ_1^n-norm which solves the optimization problem

$$\arg \min_{Sz = \mathscr{M}} \|z\|_{\ell_1^n}. \tag{2.13}$$

In this context, the updates for $X^{\ell+1}$ and $W^{\ell+1}$ in the IRLS-M algorithm reduce to vector updates $x^{\ell+1} \in \mathbb{C}^n$ and $w^{\ell+1} \in \mathbb{R}_+^n$.

Although iteratively reweighted least squares algorithms for solving constrained ℓ_1^n minimization problems of the form (2.13) have been around for half a century, it was not until recently [10] that conditions on the matrix $S \in M_{m \times n}$ were provided under which an iteratively reweighted least squares algorithm could be *proven* to converge to the solution of the ℓ_1^n minimization problem (2.13). In the algorithm proposed in [10], the weight vector w^ℓ is updated slightly differently than in the current setting; in [10] it is updated according to $w_i^\ell = (|x_i^\ell|^2 + \varepsilon_\ell^2)^{-1/2}$, while our algorithm uses the update rule

$$w_i^\ell = \left\{ \begin{array}{ll} |x_i^\ell|^{-1}, & |x_i^\ell| \geq \varepsilon_\ell, \\ \varepsilon_\ell^{-1}, & \text{else.} \end{array} \right.$$

Despite this difference, the assumptions, which we shall require on the measurements $\mathscr{S}$ in order to prove convergence of the IRLS-M algorithm to the nuclear-norm solution (2.6), draw upon the the variational framework introduced in [10]. However, our update rule for the weights (2.2) is reminiscent of a previously studied IRLS algorithm [8], which was introduced in the slightly different context of *total variation minimization* as opposed to ℓ_1 minimization. However, in that algorithm, the parameter $\varepsilon_\ell = \varepsilon > 0$ is fixed throughout the iteration; the corresponding iterations x_ε^ℓ are shown to converge to a solution $x = x(\varepsilon)$, and one can show that $x(\varepsilon)$ converges to the ℓ_1-norm solution x^* as $\varepsilon \to 0$ provided that x^* is unique. For an overview of iteratively reweighted least squares minimization, we refer the reader to [10].

2.3. Convergence results. As a consequence of the main result of this paper, Theorem 6.11, we have the following convergence result for the IRLS-algorithm under the assumption that the measurement map has the restricted isometry property in Definition 1.1.

PROPOSITION 2.1. *Consider the IRLS-M algorithm with parameters $\gamma = 1/n$ and $K \in \mathbb{N}$. Let $\mathscr{S} : M_{n \times p} \to \mathbb{C}^m$ be a surjective map with restricted isometry constants δ_{3K}, δ_{4K} satisfying*

$\eta = \frac{\sqrt{2}\delta_{4K}}{1-\delta_{3K}} < 1 - \frac{2}{K-2}$. *Then, if there exists a k-rank matrix X satisfying $\mathscr{S}(X) = \mathscr{M}$ with* $k < K - \frac{2\eta}{1-\eta}$, *the sequence $(X^\ell)_{\ell \in \mathbb{N}}$ converges to X.*

Actually, Theorem 6.11 proves something stronger: the IRLS-M algorithm is *robust*, in the sense that under the same conditions on the measurement map $\mathscr{S}$, the accumulation points of the IRLS-M algorithm are guaranteed to approximate an arbitrary $X \in M_{n\times p}$ from the measurements $\mathscr{M} = \mathscr{S}(X)$ to within a factor of the *best k-rank approximation error* of X in the nuclear norm.

The IRLS-M algorithm also has nice features independent of the affine measurements. In particular, the parameter ε_ℓ can serve as an *a posteriori* check on whether or not we have converged to a low-rank (or approximately low-rank) matrix consistent with the observations:

REMARK 2.2. *If $\varepsilon_{\ell-1} > \varepsilon$ and $\varepsilon_\ell = \gamma\sigma_{K+1}(X^\ell) = \varepsilon$, then $\|X^\ell - X^\ell_{[K]}\|_{op} \leq \varepsilon/\gamma$. In particular, if $\varepsilon_\ell = 0$, then X^ℓ is a K-rank matrix.*

3. Numerical Experiments. In the matrix completion setup (1.2) where the affine measurements are entries of the underlying matrix, the restricted isometry property fails, and so Proposition 2.1 does not apply. Nevertheless, we illustrate in this section numerical experiments which show that the IRLS-M algorithm still works very well in this setting for recovering low-rank matrices, and is competitive with state-of-the-art algorithms in the matrix completion setup; namely, the Fixed Point Continuation algorithm (FPCA), as introduced in [15], and the OptSpace algorithm, as introduced in [20]. For further numerical experiments comparing an algorithm closely related to IRLS-M with other matrix completion algorithms such as Singular Value Thresholding [4], we refer the reader to [26].
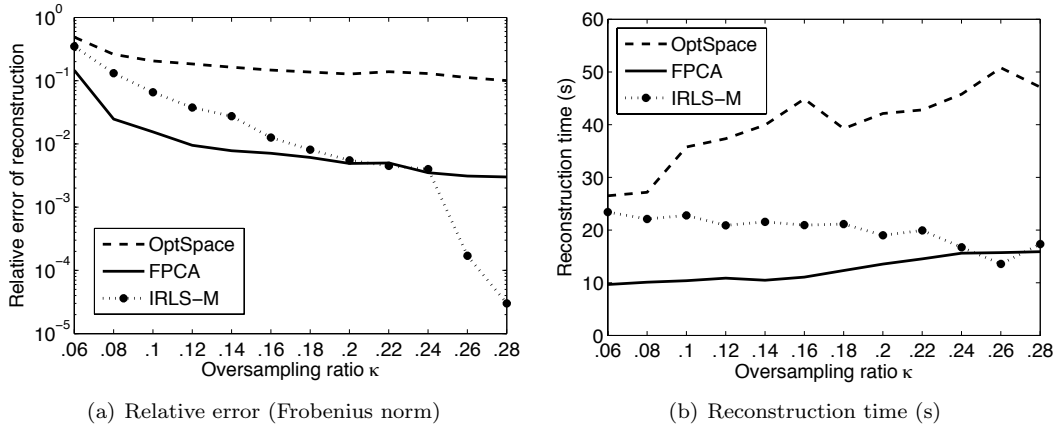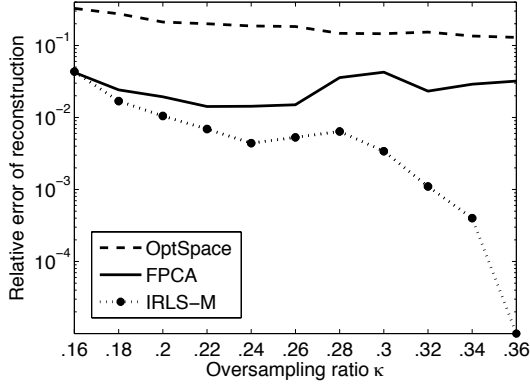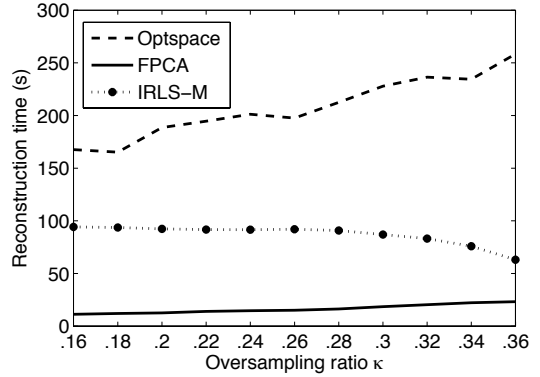

(a) Relative error (Frobenius norm)   (b) Reconstruction time (s)

FIG. 3.1. *Comparison of IRLS-M, FPCA, and OptSpace over a class of randomly-generated rank-10 matrices.*

1. In Figure 3.1, we generate rank $k = 10$ matrices X of dimensions $n = p = 500$, and of the form $X = U\Sigma V^*$, where the rows of $U \in M_{k\times n}$ and of $V \in M_{k\times n}$ are drawn independently from the multivariate standard normal distribution, and the elements of the diagonal matrix Σ are chosen independently from a standard normal distribution (note that U, Σ, and V do not comprise a singular value decomposition, although this process does generate matrices of rank $k = 10$). We plot out of 50 trials (a) the average relative error $\|X - \widetilde{X}\|_F/\|X\|_F$ and (b) the average running time in reconstructing such matrices from $m = \kappa np$ randomly chosen observed entries (matrix completion). To produce Figure 3.2, we repeat the same procedure, except with the rank of the underlying matrices now fixed to be $k = 30$.
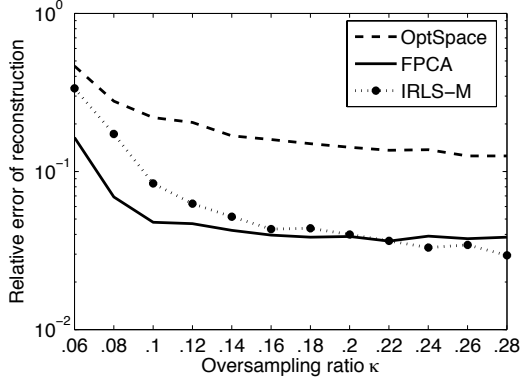
7

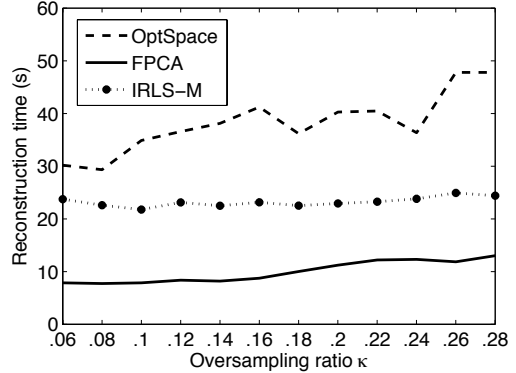(a) Relative error (with respect to Frobenius norm)   (b) Reconstruction time (s)

Fig. 3.2. *Comparison of IRLS-M, FPCA, and OptSpace over a class of randomly-generated rank-30 matrices.*

2. Figures 3.3 and 3.4 are analogs of Figures 3.1 and 3.2 respectively, generated by matrices X of rank $k = 10$ and 30 and of dimension $n = p = 500$ using the same model. But now the observed entries $\bar{X}_{i,j} = X_{i,j} + .1\mathcal{N}_{i,j}$ are corrupted by additive noise; the $\mathcal{N}_{i,j}$ are independent and identically distributed standard normal random variables.

(a) Relative error (with respect to Frobenius norm)   (b) Reconstruction time (s)

Fig. 3.3. *Comparison of IRLS-M, FPCA, and OptSpace over a class of randomly-generated rank-10 matrices subject to measurement noise.*

3. In Figures 3.5 and 3.6 we use IRLS-M, FPCA, and OptSpace to reconstruct grayscale images from partial pixel measurements. We are not suggesting that matrix completion algorithms should be used for image inpainting, but that we use images as a more realistic (nonrandom) set of matrices which are well-approximated by low-rank matrices. In general, grayscale images correspond to matrices with sharply-decaying singular values. In all experiments, the observed pixel measurements are chosen independently from the uniform distribution. The displayed images were constructed by running each algorithm using the partial pixel measurements, and then thresholding all negative pixel values to be zero.

Our results indicate that IRLS-M outperforms OptSpace for reconstructing low-rank and
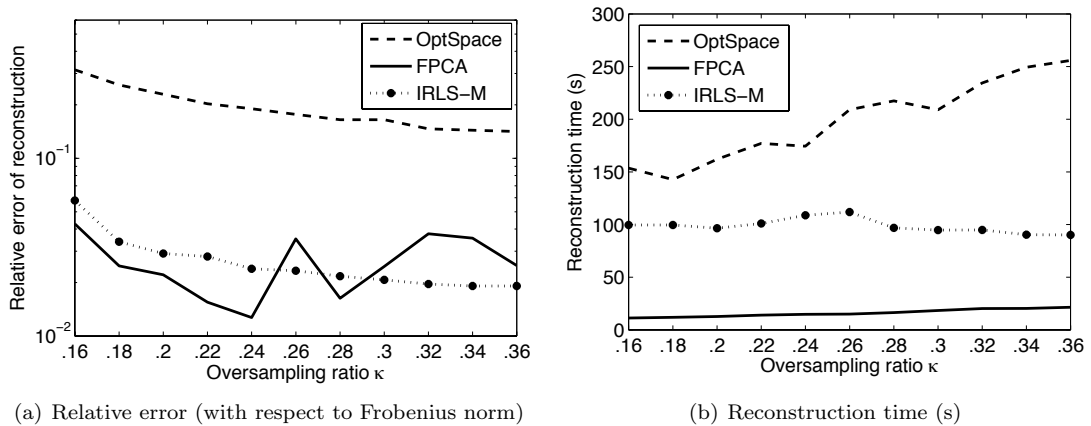
(a) Relative error (with respect to Frobenius norm)   (b) Reconstruction time (s)

FIG. 3.4. *Comparison of IRLS-M, FPCA, and OptSpace over a class of randomly-generated rank-30 matrices subject to measurement noise.*

(a) Original image   (b) Best rank-16 approximation   (c) 50% observed pixels

(d) IRLS-M reconstruction   (e) FPCA reconstruction   (f) OptSpace reconstruction
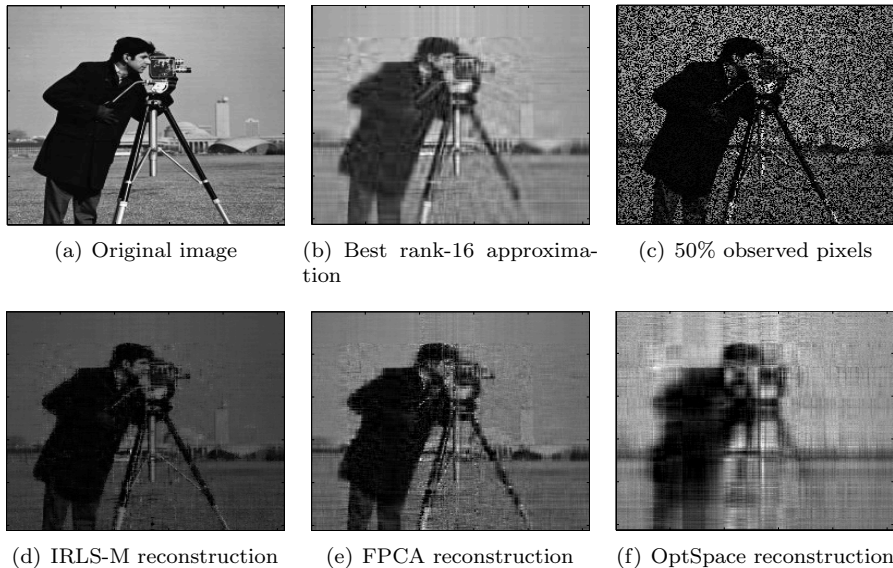
FIG. 3.5. *Figure 3.5(a) is the original 256 × 256 Camera man image. Figure 3.5(b) is its best rank-16 approximation. Figure 3.5(c) displays the 50% randomly-distributed partial pixel measurements. Figures 3.5(d), 3.5(e), and 3.5(f) are the result of applying IRLS-M, FPCA, and OptSpace, respectively, using these partial measurements and with input rank r = 16, followed by thresholding all negative pixel values to zero. The relative error (with respect to the Frobenius norm) of the best rank-16 approximation, IRLS-M, FPCA, and OptSpace, respectively are .1302, .1270, .1646, and .2157.*

approximately low-rank matrices whose singular values exhibit fast decay. Our results indicate that IRLS-M is competitive with FPCA in accuracy and superior to FPCA in accuracy once the number of measurements is sufficiently high. Nevertheless, IRLS-M can be slower in speed than FPCA. Figures 3.5 and 3.6 suggest that IRLS-M has the potential to reconstruct approximately low-rank matrices whose singular values exhibit sharp decay more accurately than FPCA.

REMARK 3.1. Much of the literature on matrix completion emphasizes the difference
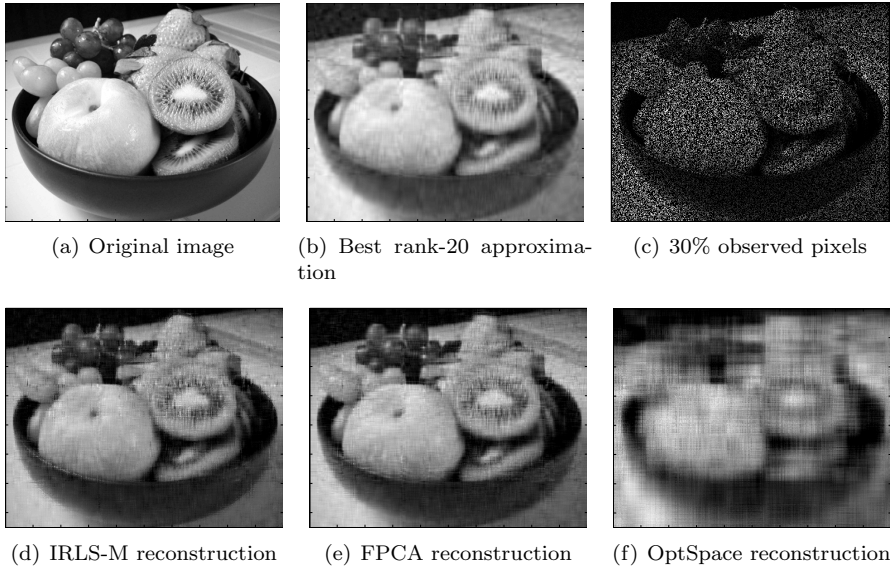
(a) Original image   (b) Best rank-20 approxima-tion   (c) 30% observed pixels

(d) IRLS-M reconstruction   (e) FPCA reconstruction   (f) OptSpace reconstruction

FIG. 3.6. *Figure 3.6(a) is the original 375×500 Fruit Bowl image. Figure 3.6(b) is its best rank-20 approximation. Figure 3.6(c) displays the 30% randomly-distributed partial pixel measurements. Figures 3.6(d), 3.6(e), and 3.6(f) are the result of applying IRLS-M, FPCA, and OptSpace, respectively, using these partial measurements with input rank $r = 20$, followed by thresholding all negative pixel values to zero. The relative error (with respect to the Frobenius norm) of the best rank-20 approximation, IRLS-M, FPCA, and OptSpace, respectively are .0955, .1066, .1131, and .2073.*

between "hard" matrix completion problems and "easy" matrix completion problems, as determined by the parameters n, p, k, and m; generally speaking, "hard" problems correspond to the situation where the number of measurements m is close to the theoretical lower bound $m = Ck \max\{n, p\} \log^2(\max\{n, p\})$. For more details, we refer the reader to [20]. We note that the range of parameters $n = p, m$, and k considered in Figure 3.1 and Figure 3.2 corresponds to a transition from "hard" to "easy" problems.

REMARK 3.2. In all above numerical experiments, we ran the IRLS-M algorithm with parameter $\gamma = 1$ as this greedy choice gives the best speed of convergence in practice, even though we could only prove convergence of the algorithm for $\gamma = 1/n$ as in Proposition 2.1. We stopped the IRLS-M iterations when either 200 iterations have been reached or $|\varepsilon_\ell - \varepsilon_{\ell-1}|/\varepsilon_\ell \le 10^{-6}$ for more than 50 consecutive iterations ℓ. These are the default parameters on the Matlab code we have made available online.[2] (One part of the code is implemented in C because Matlab prohibited fast implementation of the sequence of problems in(4.5) below.)

We ran the OptSpace algorithm using the publicly-available Matlab code by the authors[3] and using the default parameters provided therein. Finally, we ran the FPCA algorithm using publicly-available Matlab code by the authors[4]. We found that the running time of this algorithm improved considerably when we changed the settings to the "easy" problem settings, even when the default setting was "hard", and we report the performance of FPCA with this modification.

REMARK 3.3. All three algorithms are "rank-aware", meaning that the underlying rank r

---

[2] http://www.cims.nyu.edu/∼rward/IRLSM.zip
[3] http://www.stanford.edu/∼raghuram/optspace/code.html
[4] http://www.columbia.edu/∼sm2756/FPCA.htm

is provided as input. In practice most matrices of interest are only approximately low-rank, and the best choice of r is not known a priori. In the image reconstruction examples in Figures 3.5 and 3.6, we picked a rank k and number of measurements m for each image from a collection of several different attempts based on visual inspection. More generally, a systematic way to select the rank is using *cross-validation*, see [36] for details.

4. Computational Considerations.

4.1. Updating W. In order to perform the weight update,

$$W^\ell = U^\ell (\Sigma_{\varepsilon_\ell}^\ell)^{-1} (U^\ell)^*,$$

one does not need to compute all n singular vectors of W^ℓ, but rather only the first r, where r is the largest integer for which $\sigma_r > \varepsilon_\ell/\gamma$. This is because W^ℓ decomposes as the sum of a rank r matrix and a perturbation of the identity. In practice, it is sufficient to take $\gamma = 1$ (this was done in the numerical experiments of the last section); in this case, r is on the order of K, and W^ℓ decomposes into the sum of a low-rank matrix and a perturbation of the identity.

4.2. Updating X. Consider now the matrix update,

$$\bar{X} = \arg \min_{\mathscr{S}(X)=\mathscr{M}} \mathcal{J}(X,W).$$

In the matrix completion setting, the operator $\mathscr{S}$ is *separable*, i.e., acts columnwise,

$$\mathscr{S}(X) = (\mathscr{S}_1 X_1, \dots, \mathscr{S}_p X_p) = (\mathscr{M}_1, \dots, \mathscr{M}_p) = \mathscr{M} \in \mathbb{C}^m,$$

where X_i, $i = 1, \dots, p$, are the columns of X, and $\mathscr{S}_i$, $i = 1, \dots, p$ are suitable matrices acting on the vectors X_i. In this case $\|W^{1/2}X\|_F^2 = \sum_{i=1}^p \|W^{1/2}X_i\|_{\ell_2^n}^2$, so that

$$\bar{X} = \arg \min_{\mathscr{S}(X)=\mathscr{M}} \mathcal{J}(X,W) \Leftrightarrow \bar{X}_i = \arg \min_{\mathscr{S}_i X_i=\mathscr{M}_i} \|W^{1/2}X_i\|_{\ell_2^n}^2, \quad i = 1, \dots, p.$$

Note that

$$\bar{X}_i = W^{-1}\mathscr{S}_i^*(\mathscr{S}_i W^{-1}\mathscr{S}_i^*)^{-1}\mathscr{M}_i, \quad i = 1, \dots, p, \tag{4.1}$$

where the actions here are simply matrix multiplications. To speed up the computation of $(\mathscr{S}_i W^{-1}\mathscr{S}_i^*)^{-1}\mathscr{M}_j$, one may easily parallelize the independent column updates (4.1).

4.2.1. The Woodbury matrix identity. The updates in (4.1) can be further expedited by exploiting the *Woodbury matrix identity*, first introduced in [38]. It states that for dimensions $r \le n$, if $A \in M_{n\times n}$, $U \in M_{n\times r}$, $C \in M_{r\times r}$, and $V \in M_{n\times r}$, then

$$(A + UCV^*)^{-1} = A^{-1} - A^{-1}U(C^{-1} + V^*A^{-1}U)^{-1}V^*A^{-1}, \tag{4.2}$$

assuming all of the inverses exist.

We may write the inverse weight matrix W^{-1} as the sum of a diagonal rank-r matrix and a perturbation of the identity:

$$W^{-1} = \varepsilon I + U\widetilde{\Sigma}_{[r]}U^*$$

where $\widetilde{\Sigma}_{[r]} = (\Sigma - \varepsilon I)_{[r]} = \mathrm{diag}(\max\{\sigma_j - \varepsilon, 0\})$. If $\gamma = 1$ is set in the algorithm, then W^{-1} decomposes into the sum of a K-rank matrix and a perturbation of the identity during iterations ℓ where $\varepsilon_\ell < \varepsilon_{\ell-1}$. The matrix $\mathscr{S}_i W^{-1}\mathscr{S}_i^*$ may then be expanded as

$$\mathscr{S}_i W^{-1}\mathscr{S}_i^* = \varepsilon\mathscr{S}_i\mathscr{S}_i^* + (\mathscr{S}_i U)\widetilde{\Sigma}_{[r]}(\mathscr{S}_i U)^*. \tag{4.3}$$

Recall that $\mathscr{S}_i \in M_{m_i \times n}$, and that $\sum_{j=1}^{p} m_i = m$. Setting $Q_i = \mathscr{S}_i \mathscr{S}_i^*$ and $R_i = \mathscr{S}_i U$, and applying the Woodbury matrix identity,

$$(\mathscr{S}_i W^{-1} \mathscr{S}_i^*)^{-1} = \varepsilon^{-1} Q_i^{-1} \Big[I - R_i \Big(\varepsilon \widetilde{\Sigma}_{[r]}^{-1} + R_i^* Q_i^{-1} R_i \Big)^{-1} R_i^* Q_i^{-1} \Big]. \tag{4.4}$$

As $\widetilde{\Sigma}_{[r]}$ is a diagonal matrix, its inverse is trivially computed. In the setting of matrix completion, the operator $\mathscr{S}_i$ is simply a subset of rows from the $n \times n$ identity matrix, and $\mathscr{S}_i \mathscr{S}_i^T = I \in M_{m_i \times m_i}$, and $U_i = R_i = \mathscr{S}_i U \in M_{m_i \times r}$ is a subset of m_i rows from the matrix U. Therefore, utilizing the formula (4.4), this reduces to

$$(\mathscr{S}_i W^{-1} \mathscr{S}_i^*)^{-1} = \varepsilon^{-1} \Big[I - U_i \big(\varepsilon \widetilde{\Sigma}_{[r]}^{-1} + U_i^* U_i \big)^{-1} U_i^* \Big], \tag{4.5}$$

and one essentially needs to solve a linear system corresponding to the matrix $\varepsilon \widetilde{\Sigma}_{[r]}^{-1} + U_i^* U_i \in M_{r \times r}$ as opposed to a system involving the inversion of a matrix in $M_{m_i \times m_i}$. When $\gamma = 1$ and $r = K + 1$, such as at iterations ℓ where $\varepsilon_\ell < \varepsilon_{\ell-1}$, the average number of measurement entries in any particular column is at least $\widehat{m} = O(k \log np) \gg k$, so this amounts to a great reduction in computational cost.

5. Variational Interpretation.

5.1. A functional of matrices and its derivatives. In this section we show that the low-rank matrix recovery can be reformulated as an alternating minimization of a functional of matrices. Assume $X \in M_{n \times p}$, $0 \prec W = W^* \in M_{n \times n}$, and consider the functional,

$$\mathcal{J}(X, W) := \frac{1}{2} \left(\|W^{1/2} X\|_F^2 + \|W^{-1/2}\|_F^2 \right). \tag{5.1}$$

In the following subsections we shall prove that the IRLS-M algorithm may be reformulated as follows.

IRLS-M algorithm for low-rank matrix recovery: Initialize by taking $W^0 := I \in M_{n \times n}$. Set $\varepsilon_0 := 1$, $K \geq k$, and $\gamma > 0$. Then recursively define, for $\ell = 1, 2, \ldots$,

$$X^\ell := \arg \min_{\mathscr{S}(X) = \mathscr{M}} \mathcal{J}(X, W^\ell),$$

$$\varepsilon_\ell := \min \left\{ \varepsilon_{\ell-1}, \gamma \sigma_{K+1}(X^\ell) \right\}. \tag{5.2}$$

The up-date of the weight matrix W^ℓ follows a variational principle as well, i.e.,

$$W^\ell := \arg \min_{0 \prec W = W^* \preceq \varepsilon_\ell^{-1} I} \mathcal{J}(X^\ell, W) \tag{5.3}$$

The algorithm stops if $\varepsilon_\ell = 0$; in this case, define $X^j := X^\ell$ for $j > \ell$. In general, the algorithm generates an infinite sequence $(X^\ell)_{\ell \in \mathbb{N}}$ of matrices.

5.1.1. Optimization of $\mathcal{J}$ with respect to X. Let us address the matrix optimization problems and explicitly compute their solutions. For consistency of notation, we need to introduce the following *left-multiplier operator* $\mathcal{W}^{-1} : M_{n \times p} \to M_{n \times p}$ defined by $\mathcal{W}^{-1}(X) := W^{-1} X$. With this notation we can write explicitly the solution of the minimization of $\mathcal{J}$ with respect to X.

LEMMA 5.1. *Assume that $W \in M_{n \times n}$ and $W = W^* \succ 0$. Then the minimizer*

$$\bar{X} = \arg \min_{\mathscr{S}(X) = \mathscr{M}} \mathcal{J}(X, W)$$

12

is given by

$$\bar{X} = \left\{ \mathcal{W}^{-1} \circ \mathscr{S}^* \circ [\mathscr{S} \circ \mathcal{W}^{-1} \circ \mathscr{S}^*]^{-1} \right\} (\mathcal{M}).$$

Proof. For $W = W^* \succ 0$, we have

$$\bar{X} = \arg \min_{\mathscr{S}(X) = \mathcal{M}} \|W^{1/2} X\|_F^2$$

if and only if there exists a $\lambda \in \mathbb{C}^m$ such that

$$\mathscr{S}^*(\lambda) = W\bar{X}, \quad \text{or } \langle W\bar{X}, H \rangle = 0, \text{ for all } H \in \ker \mathscr{S}. \tag{5.4}$$

This equivalence can be shown directly: assume $\bar{X}$ optimal and $H \in \ker \mathscr{S}$, then

$$\|W^{1/2}\bar{X}\|_F^2 = \mathrm{Tr}(W\bar{X}\bar{X}^*) \leq \mathrm{Tr}(W(\bar{X} + H)(\bar{X} + H)^*)$$
$$= \mathrm{Tr}(W\bar{X}\bar{X}^*) + \mathrm{Tr}(WHH^*) + \mathrm{Tr}[W(H\bar{X}^* + \bar{X}H^*)].$$

Actually $\bar{X}$ is optimal if and only if $\mathrm{Tr}[W(H\bar{X}^* + \bar{X}H^*)] = 0$ for all $H \in \ker \mathscr{S}$. By the properties of the trace we have

$$\mathrm{Tr}(WH\bar{X}^*) = \overline{\mathrm{Tr}[(WH\bar{X}^*)^*]} = \overline{\mathrm{Tr}(\bar{X}H^*W)} = \overline{\mathrm{Tr}(W\bar{X}H^*)} = \overline{\mathrm{Tr}(H^*W\bar{X})}.$$

Hence $\mathrm{Tr}[W(H\bar{X}^* + \bar{X}H^*)] = 0$ if and only if $\Re e(\langle W\bar{X}, H \rangle) = 0$. This observation allows us to compute $\bar{X}$ explicitly. If we assume for a moment that actually $\langle W\bar{X}, H \rangle = 0$ for all $H \in \ker \mathscr{S}$, then necessarily $W\bar{X} = \mathscr{S}^*(\lambda)$, and we obtain $\bar{X} = \mathcal{W}^{-1} \circ \mathscr{S}^*(\lambda)$. Hence $\mathcal{M} = \mathscr{S}(\bar{X}) = \mathscr{S} \circ \mathcal{W}^{-1} \circ \mathscr{S}^*(\lambda)$ and $\lambda = [\mathscr{S} \circ \mathcal{W}^{-1} \circ \mathscr{S}^*]^{-1}(\mathcal{M})$. Eventually $\bar{X} = \left\{ \mathcal{W}^{-1} \circ \mathscr{S}^* \circ [\mathscr{S} \circ \mathcal{W}^{-1} \circ \mathscr{S}^*]^{-1} \right\}(\mathcal{M})$ satisfies by construction $\langle W\bar{X}, H \rangle = 0$, hence $\Re e(\langle W\bar{X}, H \rangle) = 0$, and the optimality condition holds. □

5.1.2. Optimization of $\mathcal{J}$ with respect to W. In this section we address the solution of the following optimization problem

$$\bar{W} = \arg \min_{0 \prec W = W^* \preceq \varepsilon^{-1} I} \mathcal{J}(X, W).$$

Let us collect a few preliminary lemmas in this direction.

LEMMA 5.2. *Under the constraint $W = W^* \succ 0$, the Fréchet derivative of $\mathcal{J}$ with respect to its second variable is*

$$\partial_W \mathcal{J}(X, W) = XX^* - W^{-2}. \tag{5.5}$$

Proof. Let us consider the Gâteaux differentiation with respect to $V = V^*$,

$$\frac{\partial}{\partial V} \|W^{1/2} X\|_F^2 := \lim_{h \to 0} \mathrm{Tr}\left[\left(\frac{(W + hV) - W}{h} \right) XX^* \right]$$
$$= \mathrm{Tr}(VXX^*) = \langle XX^*, V \rangle.$$

Since the self-adjoint and positive-definite matrices form an open set, we have

$$\partial_W \|W^{1/2} X\|_F^2 = XX^*. \tag{5.6}$$

It remains to address the Fréchet derivative of $\|W^{-1/2}\|_F^2 = \text{Tr}(W^{-1})$. We have

$$\frac{\partial}{\partial V}\|W^{-1/2}\|_F^2 = \lim_{h\to 0} h^{-1}\left\{\text{Tr}[(W+hV)^{-1}] - \text{Tr}(W^{-1})\right\}$$

$$= \text{Tr}\left[\lim_{h\to 0}\frac{(W+hV)^{-1} - W^{-1}}{h}\right].$$

Note that $(W+hV)\left[\frac{(W+hV)^{-1}-W^{-1}}{h}\right] = -VW^{-1}$, so

$$\lim_{h\to 0}\frac{(W+hV)^{-1} - W^{-1}}{h} = -W^{-1}VW^{-1}.$$

By using the cyclicity of the trace we obtain $\partial_W\|W^{-1/2}\|_F^2 = \langle W^{-2}, V\rangle$, and therefore

$$\frac{\partial}{\partial W}\|W^{-1/2}\|_F^2 = -W^{-2}. \tag{5.7}$$

This latter statement can also be obtained by applying the more general result [24, Proposition 6.2]. Putting together (5.6) and (5.7) yields (5.5). □

Using a similar argument as in Lemma 5.2, we arrive at the next result.

PROPOSITION 5.3. *Consider $X = U\Sigma V^*$, where $\Sigma = \text{diag}(\sigma_1, ..., \sigma_n)$. Let $\varepsilon > 0$. Then*

$$\arg\min_{0\prec W = W^* \preceq \varepsilon^{-1}I} \mathcal{J}(X, W) =: \bar{W} = U\Sigma_\varepsilon^{-1}U^*. \tag{5.8}$$

We recall that $\Sigma_\varepsilon = \text{diag}(\max\{\sigma_j, \varepsilon\})$.

Proof. First observe that $\mathcal{J}(X, W) = \infty$ if W is not invertible so that

$$\bar{W} = \arg\min_{0\preceq W = W^* \preceq \varepsilon^{-1}I} \mathcal{J}(X, W).$$

That is, the constraint of positive-definiteness can be relaxed to positive semi-definiteness. Now, the set $\mathcal{C}_0 = \{W \in M_{n\times n} : 0 \preceq W = W^* \preceq \varepsilon^{-1}I\}$ is a convex set, and may be reformulated as

$$\mathcal{C}_0 = \{W \in M_{n\times n} : 0 \preceq W = W^*, \ \|W\|_{op} \le \varepsilon^{-1}\}$$
$$= \{W \in M_{n\times n} : 0 \preceq W = W^*, \ \sup_{\|u\|_{\ell_2^n}=1}\|W^{1/2}u\|_{\ell_2^n}^2 = \sup_{\|u\|_{\ell_2^n}=1}\|Wu\|_{\ell_2^n} \le \varepsilon^{-1}\}.$$

Let $e_i \in \mathbb{R}^n$ denote the ith canonical basis vector, and note that $\|Ue_i\|_{\ell_2^n} = 1$. Then it is clear that $\mathcal{C}_0 \subset \mathcal{C}_1$, where $\mathcal{C}_1$ is the convex set

$$\mathcal{C}_1 = \{W \in M_{n\times n} : 0 \preceq W = W^*, \ \|W^{1/2}Ue_i\|_{\ell_2^n}^2 \le \varepsilon^{-1}, \quad i = 1, ..., n\}.$$

The Lagrangian corresponding to the optimization of $\mathcal{J}$ over the convex set $\mathcal{C}_1$ is given by

$$\mathcal{L}(X, W, \lambda) := \frac{1}{2}\left[\|W^{1/2}X\|_F^2 + \|W^{-1/2}\|_F^2 + \sum_{i=1}^n \lambda_i\left(\|W^{1/2}Ue_i\|_{\ell_2^n}^2 - \varepsilon^{-1}\right)\right];$$

by weak duality [3, p. 226] the minimum of $\mathcal{J}(X, \cdot)$ over $\mathcal{C}_1$ can be bounded from below by the Lagrangian dual function,

$$\sup_{(\lambda_1,...,\lambda_n)\ge 0} g(\lambda) \le \min_{W\in\mathcal{C}_1} \mathcal{J}(X, W), \qquad g(\lambda) := \min_{W\in\mathcal{C}_1} \mathcal{L}(X, W, \lambda).$$

14

Following an argument similar to Lemma 5.2, $\bar{W}_\lambda = \arg\min_{W \in \mathcal{C}_1} \mathcal{L}(X, W, \lambda)$ agrees with the solution to the Karush-Kuhn-Tucker conditions,

$$XX^* + \sum_{i=1}^n \lambda_i U E_{ii} U^* = W^{-2}, \tag{5.9}$$

or

$$W = U \operatorname{diag}(\sigma_i^2 + \lambda_i)^{-1/2} U^*,$$

if $W \in \mathcal{C}_1$, or equivalently if $\lambda_i \geq \max\{\varepsilon^2 - \sigma_i^2, 0\}$. A straightforward calculation shows that $\bar{\lambda} = \arg\max_{\lambda_i \geq \max\{\varepsilon^2 - \sigma_i^2, 0\}} g(\lambda)$ satisfies

$$\bar{\lambda}_i = \begin{cases} \varepsilon^2 - \sigma_i^2, & \varepsilon^2 - \sigma_i^2 \geq 0, \\ 0, & \text{else}, \end{cases}$$

while $g(\bar{\lambda}) = \frac{1}{2} \sum_{\sigma_i \leq \varepsilon} \left(\frac{\varepsilon^2 + \sigma_i^2}{\varepsilon}\right) + \sum_{\sigma_i \geq \varepsilon} \sigma_i$. But at $\bar{W}_{\bar{\lambda}} = U \operatorname{diag}(\sigma_i^2 + \bar{\lambda}_i)^{-1/2} U^*$, we have the equality $g(\bar{\lambda}) = \mathcal{J}(\bar{W}_{\bar{\lambda}}, X)$. By duality, $\bar{W}_{\bar{\lambda}}$ must realize the minimal value of $\mathcal{J}(X, \cdot)$ over the set $\mathcal{C}_1$. As $\bar{W}_{\bar{\lambda}}$ is also contained in $\mathcal{C}_0$ and as $\mathcal{C}_0 \subset \mathcal{C}_1$, it follows that $\bar{W}_{\bar{\lambda}}$ also minimizes $\mathcal{J}(X, \cdot)$ over the original convex set $\mathcal{C}_0$. We set $\bar{W} = \bar{W}_{\bar{\lambda}}$. □

6. Analysis of the Algorithm.

6.1. Basic properties of the iterates. We start with simple properties of the iterates that hold without any requirements. In order to show the actual convergence of the IRLS-M algorithm along with the relationship to nuclear norm minimization, we will assume additional properties of the measurement map $\mathscr{S}$ in the next subsection.

PROPOSITION 6.1. *Assume that $(X^\ell, W^\ell, \varepsilon_\ell)_{\ell \in \mathbb{N}}$ is the output of the main algorithm. Then the following properties hold.*

(i) $\mathcal{J}(X^{\ell+1}, W^{\ell+1}) \leq \mathcal{J}(X^\ell, W^\ell)$, for all $\ell \in \mathbb{N}$;

(ii) $\mathcal{J}(X^\ell, W^\ell) \geq \|X^\ell\|_*$, for all $\ell \in \mathbb{N}$;

(iii) There exists a constant $\mathscr{A} > 0$ such that

$$2\mathscr{A} \left[\mathcal{J}(X^\ell, W^\ell) - \mathcal{J}(X^{\ell+1}, W^{\ell+1}) \right] \geq \|X^{\ell+1} - X^\ell\|_F^2. \tag{6.1}$$

This in particular implies

$$\lim_{\ell \to \infty} \|X^{\ell+1} - X^\ell\|_F^2 = 0. \tag{6.2}$$

Proof. (i) By the optimality properties of $X^{\ell+1}$ and $W^{\ell+1}$, along with the admissibility of W^ℓ in the optimization of $W^{\ell+1}$ (as $\varepsilon_{\ell+1} \leq \varepsilon_\ell$), we obtain the following chain of inequalities

$$\mathcal{J}(X^{\ell+1}, W^{\ell+1}) \leq \mathcal{J}(X^{\ell+1}, W^\ell) \leq \mathcal{J}(X^\ell, W^\ell).$$

(ii) A direct computation shows

$$\mathcal{J}(X^\ell, W^\ell) = \left[\sum_{i=1}^r \sigma_i^\ell + \sum_{i=r+1}^n \frac{(\sigma_i^\ell)^2 + \varepsilon_\ell^2}{2\varepsilon_\ell} \right],$$

where r is the largest integer for which $\sigma_r^\ell \geq \varepsilon_\ell$. Since $(\sigma_i^\ell)^2 + \varepsilon_\ell^2 \geq 2\varepsilon_\ell \cdot \sigma_i^\ell$, we have

$$\mathcal{J}(X^\ell, W^\ell) \geq \|X^\ell\|_*.$$

15

(iii) We have the following estimate,

$$2\left[\mathcal{J}(X^\ell, W^\ell) - \mathcal{J}(X^{\ell+1}, W^{\ell+1})\right] \geq \left[\mathcal{J}(X^\ell, W^\ell) - \mathcal{J}(X^{\ell+1}, W^\ell)\right]$$
$$= \langle W^\ell X^\ell, X^\ell \rangle - \langle W^\ell X^{\ell+1}, X^{\ell+1} \rangle$$
$$= \langle W^\ell(X^\ell + X^{\ell+1}), X^\ell - X^{\ell+1} \rangle + 2i\Im m(\langle W^\ell X^\ell, X^{\ell+1} \rangle).$$

By optimality of $X^{\ell+1}$ and in virtue of (5.4), we can add $0 = -2\langle W^\ell X^{\ell+1}, X^\ell - X^{\ell+1} \rangle$ and $0 = 2i\Im m(\langle W^\ell(X^{\ell+1} - X^\ell), X^{\ell+1} \rangle)$ (note also that $X^\ell - X^{\ell+1} \in \ker \mathscr{S}$), and we obtain

$$2\left[\mathcal{J}(X^\ell, W^\ell) - \mathcal{J}(X^{\ell+1}, W^{\ell+1})\right] \geq \langle W^\ell(X^\ell - X^{\ell+1}), X^\ell - X^{\ell+1} \rangle$$
$$= \|(W^\ell)^{1/2}(X^\ell - X^{\ell+1})\|_F^2.$$

On the one hand,

$$\|X^\ell - X^{\ell+1}\|_F^2 \leq \|(W^\ell)^{1/2}(X^\ell - X^{\ell+1})\|_F^2 \|(W^\ell)^{-1/2}\|_F^2,$$

and

$$\|(W^\ell)^{-1/2}\|_F^2 = (n-r)\varepsilon_\ell^{1/2} + \sum_{i=1}^r (\sigma_i^\ell)^{1/2}.$$

On the other hand,

$$\sigma_i^\ell \leq \|X^\ell\|_* \leq \mathcal{J}(X^1, W^0).$$

Hence, there exists a constant $\mathscr{A} > 0$ such that $\|(W^\ell)^{-1/2}\|_F^2 \leq \mathscr{A}$, for all $\ell \in \mathbb{N}$. It follows from (6.1) that

$$\sum_{\ell=0}^N \|X^{\ell+1} - X^\ell\|_F^2 \leq 2\mathscr{A}\left[\mathcal{J}(X^0, W^0) - \mathcal{J}(X^{N+1}, W^{N+1})\right] \leq 2\mathscr{A}\,\mathcal{J}(X^0, W^0).$$

This implies that the sum on the left converges as $N \to \infty$, and (6.2) follows. $\square$

6.2. Null space property. We need to enforce additional conditions in order to obtain a simultaneous promotion of the approximation to both minimal rank and minimal nuclear norm solution. For this purpose we recall the rank null space property, which is the analog to the null space property for the vector case in compressed sensing [9, 16, 12, 32].

DEFINITION 6.2. *A map $\mathscr{S} : M_{n \times p} \to \mathbb{C}^m$ satisfies the* rank null space property (RNSP) *of order k if for all $H \in \ker \mathscr{S} \setminus \{0\}$ and all decompositions $H = H_1 + H_2$ with k-rank H_1 it holds*

$$\|H_1\|_* < \|H_2\|_*.$$

The rank null space property is closely linked to the nuclear norm minimization problem

$$\min_{X \in M_{n \times p}} \|X\|_* \quad \text{subject to} \quad \mathscr{S}(X) = \mathscr{M}. \tag{6.3}$$

The following theorem has been shown in [35, Theorem 3].

THEOREM 6.3. *Let $\mathscr{S} : M_{n \times p} \to \mathbb{C}^m$ be a linear map. Then, every k-rank $X \in M_{n \times p}$ is the unique solution of the problem (6.3) for the datum $\mathscr{M} = \mathscr{S}(X)$ if and only if $\mathscr{S}$ satisfies the RNSP of order k.*

We also need a slightly stronger condition.

DEFINITION 6.4. *A map $\mathscr{S} : M_{n\times p} \to \mathbb{C}^m$ satisfies the* strong rank null space property (SRNSP) *of order k with constant $\eta \in (0,1)$ if for all $X \in \ker \mathscr{S} \setminus \{0\}$ and all decompositions $X = X_1 + X_2$ with k-rank X_1 there exists another decomposition $X = H_1 + H_2$ such that H_1 is of rank at most $2k$, $\langle H_1, H_2 \rangle = 0$, $X_1 H_2^* = 0$, $X_1^* H_2 = 0$, and*

$$\|H_1\|_* \leq \eta \|H_2\|_*.$$

REMARK 6.5. *The above definition is equivalent to [26, Definition II.4]: Let $GL(U,V)$ be the subspace of $n \times p$ matrices of rank at most k, whose row space belongs to the span of $V \in M_{n\times k}$ and whose column space belongs to the span of $U \in M_{n\times k}$. Let $S_k = \{GL(U,V) : U \in M_{n\times k}, V \in M_{p\times k}, U^*U = V^*V = I\}$. For $T \in GL(U,V)$, we let $\mathcal{P}_T$ be the projection onto T, and $T^\perp$ denotes the orthogonal complement of T in $M_{n\times p}$. Then a map $\mathscr{S} : M_{n\times p} \to \mathbb{C}^m$ satisfies the* strong rank null space property (SRNSP) *of order k with constant $\eta \in (0,1)$ if*

$$\|\mathcal{P}_T(X)\|_* \leq \eta \|\mathcal{P}_{T^\perp}\|_*$$

for all $T \in S_k$ and $H \in \ker \mathscr{S} \setminus \{0\}$.

Further, we introduce the *best k-rank approximation error* in the nuclear norm,

$$\rho_k(X)_* := \min_{Z \in M_{n\times p}:\text{rank}(Z) \leq k} \|X - Z\|_*.$$

It is well-known that the minimum is attained at the k-spectral truncation $Z = X_{[k]}$. The SRNSP implies the following crucial inequality.

LEMMA 6.6 (Inverse triangle inequality). *Assume that $\mathscr{S} : M_{n\times p} \to \mathbb{C}^m$ satisfies the SRNSP of order k and constant $\eta \in (0,1)$. Let X, Z be matrices such that*

$$\mathscr{S}(X) = \mathscr{S}(Z).$$

Then we have the following inverse triangle inequality

$$\|X - Z\|_* \leq \frac{1+\eta}{1-\eta}\left(\|Z\|_* - \|X\|_* + 2\rho_k(X)_*\right). \tag{6.4}$$

This result was independently obtained in [26, Lemma II.5]. Inverse triangle inequalities are well-known in the context of the classical null space property for vectors [10], see also the survey paper [31].

Proof. Let $X_0 := X_{[k]}$ be the best k-rank approximation of X in the nuclear norm, and set $X_c = X - X_0$. Let $H = X - Z = X_0 + X_c - Z \in \ker \mathscr{S}$. Set $X_1 = Z - X_c$, so that $H = X_0 - X_1$. By the definition of the strong null space property, there exists a decomposition $H = H_0 + H_c$ with $\text{rank}(H_0) \leq 2k$, $\langle H_0, H_c \rangle = 0$, $X_0 H_c^* = 0$, $X_0^* H_c = 0$ and such that $\|H_0\|_* \leq \eta \|H_c\|_*$. Hence, by Lemma 6.10 recalled below, we have

$$\|X_0 - H_c\|_* = \|H_c\|_* + \|X_0\|_*$$

and, as $H_0 + X_1 = X_0 - H_c$, we have equivalently

$$\|H_c\|_* + \|X_0\|_* = \|H_0 + X_1\|_* \leq \|H_0\|_* + \|X_1\|_*.$$

The inequality $\|H_0\|_* \leq \eta \|H_c\|_*$ yields

$$\|H_c\|_* + \|X_0\|_* \leq \|X_1\|_* + \eta \|H_c\|_*,$$

17

or, equivalently,

$$\|H_c\|_* \leq \frac{1}{1-\eta} \left(\|X_1\|_* - \|X_0\|_* \right).$$

Now, using $X_1 = Z - X_c$, we obtain

$$\|X - Z\|_* = \|H_0 + H_c\|_* \leq (1+\eta)\|H_c\|_* \leq \frac{1+\eta}{1-\eta}(\|X_1\|_* - \|X_0\|_*)$$

$$\leq \frac{1+\eta}{1-\eta}(\|Z\|_* - \|X_0\|_* + \|X_c\|_*) \leq \frac{1+\eta}{1-\eta}(\|Z\|_* - \|X - X_c\|_* + \|X_c\|_*)$$

$$\leq \frac{1+\eta}{1-\eta}(\|Z\|_* - \|X\|_* + 2\|X_c\|_*).$$

The proof is completed by noting that $\|X_c\|_* = \|X - X_0\|_* = \rho_k(X)_*$. $\square$

Note that if X is k-rank then $\rho_k(X)_* = 0$ and (6.4) reduces to

$$\|X - Z\|_* \leq \frac{1+\eta}{1-\eta} \left(\|X\|_* - \|Z\|_* \right). \tag{6.5}$$

Now we easily conclude that the SRNSP implies stable low-rank recovery via nuclear norm minimization.

COROLLARY 6.7. *Suppose* $\mathscr{S} : M_{n \times p} \to \mathbb{C}^m$ *satisfies the strong null space property of order k with some constant* $\eta \in (0,1)$. *Let* $X \in M_{n \times p}$ *and* $\mathscr{M} = \mathscr{S}(X)$. *Then a solution* $\bar{X}$ *of* (6.3) *satisfies*

$$\|X - \bar{X}\|_* \leq 2\frac{1+\eta}{1-\eta}\rho_k(X)_*.$$

In particular, every k-rank $X \in M_{n \times p}$ is the unique solution of the nuclear norm minimization problem (6.3). *Consequently, the strong rank null space property of order k with some constant* $\eta < 1$ *implies the rank null space property of order k.*

Proof. Clearly $\|\bar{X}\|_* \leq \|X\|_*$ and (6.4) yields $\|X - \bar{X}\|_* \leq 2(1+\eta)/(1-\eta)\rho_k(X)_*$. Since $\rho_k(X)_* = 0$ for all k-rank X, the latter estimate implies exact recovery of such matrices. Theorem 6.3 implies also that $\mathscr{S}$ satisfies the rank null space property of order k. $\square$

The SRNSP is somewhat difficult to analyze directly. As in the vector case [12, 32], it is implied by the (rank) restricted isometry property, Definition 1.1:

PROPOSITION 6.8. *Assume that* $\mathscr{S} : M_{n \times p} \to \mathbb{C}^m$ *has restricted isometry constant*

$$\delta_{4k} < \sqrt{2} - 1 \approx 0.41 . \tag{6.6}$$

Then $\mathscr{S}$ satisfies the SRNSP of order k with constant

$$\eta = \sqrt{2}\frac{\delta_{4k}}{1 - \delta_{3k}} \quad \in (0,1) .$$

Our proof uses [5, Lemma 3.3] and [34, Lemma 2.3], which we recall here for the reader's convenience.

LEMMA 6.9 (Lemma 3.3 in [5]). *Let* $\mathscr{S} : M_{n \times p} \to \mathbb{C}^m$ *with restricted isometry constant* $0 < \delta_k$. *Then for two matrices $X, Y \in M_{n \times p}$ with $\mathrm{rank}(X) + \mathrm{rank}(Y) \leq k$ and $\langle X, Y \rangle = 0$ it holds*

$$|\langle \mathscr{S}(X), \mathscr{S}(Y) \rangle| \leq \delta_k \|X\|_F \|Y\|_F.$$

LEMMA 6.10 (Lemma 2.3 in [34]). *Let X and Z be matrices of the same dimensions. If $XZ^* = 0$ and $X^*Z = 0$ then*

$$\|X + Z\|_* = \|X\|_* + \|Z\|_*.$$

Proof. [of Proposition 6.8] For convenience we give a simple proof, which results in the stated non-optimal constant in (6.6), along the lines of the proof of Theorem 2.6 in [32]. Let $X \in \ker \mathscr{S} \setminus \{0\}$ and X_1 be of rank at most k, so that, by setting $X_2 = X - X_1$, we have $X = X_1 + X_2$. Let

$$X_1 = U \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} V^*$$

be the (full) singular value decomposition of X_1 with Σ being a $k \times k$ diagonal matrix. Set $\hat{H} = U^* X V$ and partition

$$\hat{H} = \begin{pmatrix} \hat{H}_{11} & \hat{H}_{12} \\ \hat{H}_{21} & \hat{H}_{22} \end{pmatrix},$$

where $\hat{H}_{11} \in M_{k \times k}$. Now, set

$$H_0 = U \begin{pmatrix} \hat{H}_{11} & \hat{H}_{12} \\ \hat{H}_{21} & 0 \end{pmatrix} V^*, \quad H_c = U \begin{pmatrix} 0 & 0 \\ 0 & \hat{H}_{22} \end{pmatrix} V^*.$$

Then $\operatorname{rank}(H_0) \le 2k$, $X_1 H_c^* = 0$, $X_1^* H_c = 0$, and $\langle H_0, H_c \rangle = 0$ as desired. It remains to prove that $\|H_0\|_* \le \eta \|H_c\|_*$. Let $\hat{H}_{22}$ have singular value decomposition $\hat{H}_{22} = \tilde{U} \operatorname{diag}(\tilde{\sigma}) \tilde{V}^*$, in particular, the entries of $\tilde{\sigma}$ are in decreasing order. Now we choose an integer $\ell > 0$ and decompose the vector $\tilde{\sigma}$ into vectors $\sigma^{(j)}$, $j = 1, 2, \ldots$, of support length ℓ defined by

$$\sigma_i^{(j)} = \begin{cases} \tilde{\sigma}_i, & \text{if } \ell(j-1) < i \le \ell j, \\ 0, & \text{otherwise} . \end{cases}$$

Clearly, $\sum_j \sigma^{(j)} = \tilde{\sigma}$. We set, for $j = 1, 2, \ldots$,

$$H_j = U \begin{pmatrix} 0 & 0 \\ 0 & \tilde{U} \operatorname{diag}(\sigma^{(j)} \tilde{V}^*) \end{pmatrix} V^*.$$

Then $\operatorname{rank}(H_j) \le \ell$, $H_c = \sum_{j \ge 1} H_j$ and $\langle H_0, H_j \rangle = 0$ for all $j = 1, 2, \ldots$. Also, observe that since $H \in \ker \mathscr{S}$ we have $\mathscr{S}(H_0 + H_1) = \sum_{j \ge 2} \mathscr{S}(-H_j)$. Now we first use that $\|H_0\|_* \le \sqrt{2k} \|H_0\|_F \le \sqrt{2k} \|H_0 + H_1\|$ by orthogonality of H_0 and H_1, and estimate the latter. To this end, we use orthogonality of H_0 and H_1, the RIP and Lemma 6.9,

$$\|H_0 + H_1\|_F^2 \le \frac{1}{1 - \delta_{2k+\ell}} \|\mathscr{S}(H_0 + H_1)\|_F^2 = \frac{1}{1 - \delta_{2k+\ell}} \langle \mathscr{S}(H_0 + H_1), \sum_{j \ge 2} \mathscr{S}(H_j) \rangle$$

$$= \frac{1}{1 - \delta_{2k+\ell}} \sum_{j \ge 2} \langle \mathscr{S}(H_0 + H_1), \mathscr{S}(H_j) \rangle \le \frac{\delta_{2k+2\ell}}{1 - \delta_{2k+\ell}} \|H_0 + H_1\|_F \sum_{j \ge 2} \|H_j\|_F.$$

Hence, $\|H_0 + H_1\|_F \le \frac{\delta_{2k+2\ell}}{1 - \delta_{2k+\ell}} \sum_{j \ge 2} \|H_j\|_F$. Since the sequence $\tilde{\sigma}$ is nonincreasing, we have by definition of $\sigma^{(j)}$ that, for $j \ge 2$ and for all i, $\sigma_i^{(j+1)} \le \frac{1}{\ell} \|\sigma^{(j)}\|_{\ell_1} = \frac{1}{\ell} \|H_j\|_*$, and therefore,

$$\|H_{j+1}\|_F = \left(\sum_{i=j\ell+1}^{(j+1)\ell} \sigma_i^2 \right)^{1/2} \le \frac{1}{\sqrt{\ell}} \|H_j\|_*.$$

19

Combining our estimates we obtain

$$\|H_0\|_* \leq \sqrt{2k}\|H_0 + H_1\|_F \leq \frac{\delta_{2k+2\ell}}{1 - \delta_{2k+\ell}} \sqrt{2k} \sum_{j \geq 2} \|H_j\|_F \leq \frac{\delta_{2k+2\ell}}{1 - \delta_{2k+\ell}} \sqrt{\frac{2k}{\ell}} \sum_{j \geq 1} \|H_j\|_*$$

$$= \frac{\delta_{2k+2\ell}}{1 - \delta_{2k+\ell}} \sqrt{\frac{2k}{\ell}} \|H_c\|_*. \tag{6.7}$$

In the last step we applied Lemma 6.10. Now we choose $\ell = k$ and obtain $\|H_0\|_* \leq \eta\|H_c\|_*$ with

$$\eta = \sqrt{2} \frac{\delta_{4k}}{1 - \delta_{3k}}.$$

Since $\delta_{3k} \leq \delta_{4k}$, we have $\eta < 1$ if $\delta_{4k} < \sqrt{2} - 1$. □

We note that different choices of ℓ in (6.7) (say $\ell = 2k$, or $\ell = \lceil k/2 \rceil$) lead to slightly different conditions. Also, we do not claim optimality of the constant $\sqrt{2} - 1$. We expect that with a more complicated proof as in [5] or [13, 14, 25], one can still improve this value. Our goal here was to rather provide a simple proof. After submission of an initial version of our paper, we became aware of [29], where an elegant relationship between vector and matrix cases is shown. In particular, properties and recovery guarantees valid for sparse vector recovery by means of ℓ_1-minimization are shown to imply corresponding properties and recovery guarantees for low-rank matrices by means of nuclear-norm minimization. It seems that together with [25] this implies slightly better guarantees than provided in Proposition 6.8.

6.3. Analysis of the algorithm when $\mathscr{S}$ satisfies the null space property. Before presenting the main result of this paper, we introduce another functional, dependent on a parameter $\varepsilon > 0$,

$$\mathcal{J}_\varepsilon(X) := J_\varepsilon(\Sigma(X)) = \sum_{i=1}^n j^\varepsilon(\sigma_i(X)), \tag{6.8}$$

where

$$j^\varepsilon(u) = \begin{cases} |u|, & |u| \geq \varepsilon, \\ \frac{u^2 + \varepsilon^2}{2\varepsilon}, & u < \varepsilon. \end{cases}$$

THEOREM 6.11. *Consider the IRLS-M algorithm for low-rank matrix recovery, with parameters $\gamma = 1/n$ and $K \in \mathbb{N}$. Let $\mathscr{S} : M_{n \times p} \to \mathbb{C}^m$ be a surjective map. Then, for each set of measurements $\mathscr{M} \in \mathbb{C}^m$, the sequence of matrices $(X^\ell)_{\ell \in \mathbb{N}}$ produced by the IRLS-M algorithm has the following properties:*

(i) Assume that $\mathscr{S}$ satisfies the strong rank null space property of order K (Definition 6.4). If $\lim_{\ell \to \infty} \varepsilon_\ell = 0$, then the sequence $(X^\ell)_{\ell \in \mathbb{N}}$ converges to a K-rank matrix $\bar{X}$ agreeing with the measurements; in this case, $\bar{X}$ is also the unique nuclear norm minimizer.

(ii) If $\lim_{\ell \to \infty} \varepsilon_\ell = \varepsilon > 0$ then every subsequence of X^ℓ has a convergent subsequence. Each accumulation point $\tilde{X}$ of $(X^\ell)_{\ell \geq 1}$ coincides with a minimizer $\bar{X}$ of the functional $\mathcal{J}_\varepsilon$ subject to $\mathscr{S}(\bar{X}) = \mathscr{M}$. In particular, if this minimizer is unique then the full sequence $(X^\ell)_{\ell \geq 1}$ converges to it.

If, in addition, $\mathscr{S}$ satisfies the strong rank null space property of order K (Definition 6.4) and constant $\eta < 1 - \frac{2}{K-2}$, then each accumulation point $\tilde{X} = \bar{X}$ of $(X^\ell)_{\ell \geq 1}$ satisfies, for any matrix X such that $\mathscr{S}(X) = \mathscr{M}$, and for $k < K - \frac{2\eta}{1-\eta}$,

$$\|X - \bar{X}\|_* \leq \Lambda \rho_k(X)_*,$$

where

$$\Lambda := \frac{4(1+\eta)^2}{(1-\eta)^2((K-k)(1-\eta)-2\eta)} + \frac{2(1+\eta)}{1-\eta}.$$

(iii) In particular, if $\mathscr{S}$ satisfies the strong rank null space property of order K with constant $\eta < 1 - \frac{2}{K-2}$, and if there exists a k-rank matrix X satisfying $\mathscr{S}(X) = \mathscr{M}$, then necessarily $\varepsilon = 0$.

Proof. Note that since $0 \le \varepsilon_{\ell+1} \le \varepsilon_\ell$, the sequence $(\varepsilon_\ell)_{\ell \in \mathbb{N}}$ always converges.

(i) If $\varepsilon_\ell = 0$ for some ℓ, then we set $\bar{X} = X^\ell$. If $\varepsilon_\ell > 0$ for all ℓ, then there exists a subsequence, such that

$$\varepsilon_{\ell_j+1} < \varepsilon_{\ell_j}, \quad j \in \mathbb{N},$$

and $\varepsilon_{\ell_j+1} = \gamma \sigma_{K+1}(X^{\ell_j+1})$. Because of Proposition 6.1 (i) and (ii), $(X^{\ell_j+1})_j$ is bounded. Hence, we can extract a further subsequence, denoted again by $(X^{\ell_j+1})_j$ for simplicity, which converges to some $\tilde{X} = \lim_{j\to\infty} X^{\ell_j}$. Since $\lim_{j\to\infty} \varepsilon_{\ell_j+1} = 0$, an application of Theorem 7.1 immediately yields $\sigma_{K+1}(\tilde{X}) = 0$, and $\tilde{X}$ is a K-rank solution of $\mathscr{S}(X) = \mathscr{M}$.

The strong rank null space property together with Corollary 6.7 implies that $\bar{X}$ is the unique nuclear norm minimizer $\bar{X}$. Next we show that the whole sequence $(X^\ell)_\ell$ converges to $\bar{X}$. We have

$$\mathcal{J}(X^\ell, W^\ell) = \left[\sum_{i=1}^{r(\ell)} \sigma_i^\ell + \sum_{i=r(\ell)+1}^{n} \frac{(\sigma_i^\ell)^2 + \varepsilon_\ell^2}{2\varepsilon_\ell} \right],$$

where we recall that $r(\ell)$ is the smallest integer for which $\sigma_i^\ell < \varepsilon_\ell$ if $i > r(\ell)$. Since $\mathcal{J}(X^\ell, W^\ell, \varepsilon_\ell)$ is a monotonically decreasing sequence and $\mathcal{J}(X^{\ell_j}, W^{\ell_j}, \varepsilon_{\ell_j}) \to \|\bar{X}\|_*$ for $j \to \infty$, we also have $\mathcal{J}(X^\ell, W^\ell, \varepsilon_\ell) \to \|\bar{X}\|_*$ for $\ell \to \infty$. We have the estimates from above and below

$$\mathcal{J}(X^\ell, W^\ell, \varepsilon_\ell) - \sum_{i=r(\ell)+1}^{n} \left(\frac{(\sigma_i^\ell)^2 + \varepsilon_\ell^2}{2\varepsilon_\ell} - \sigma_i^\ell \right) \le \|X^\ell\|_* \le \mathcal{J}(X^\ell, W^\ell, \varepsilon_\ell).$$

As $\ell \to \infty$, $\sigma_i^\ell < \varepsilon_\ell \to 0$ for $i \ge r(\ell) + 1$, and we compute

$$\lim_{\ell\to\infty} \sum_{i=r(\ell)+1}^{n} \left(\frac{(\sigma_i^\ell)^2 + \varepsilon_\ell^2}{2\varepsilon_\ell} - \sigma_i^\ell \right) \le \lim_{\ell\to\infty} \sum_{i=r(\ell)+1}^{n} \left(\frac{\varepsilon_\ell - \sigma_i^\ell}{2} \right) \le \lim_{\ell\to\infty} n\varepsilon_\ell = 0,$$

and we obtain

$$\lim_{\ell\to\infty} \|X^\ell\|_* = \|\bar{X}\|_*.$$

Since $\bar{X}$ is a K-rank matrix, we can apply the inverse triangle inequality (6.5) to obtain

$$\|X^\ell - \bar{X}\|_* \le \frac{1+\eta}{1-\eta} \left(\|X^\ell\|_* - \|\bar{X}\|_* \right), \tag{6.9}$$

where $\eta < 1$ by assumption. Taking the limit for $\ell \to \infty$ in (6.9), we obtain $\lim_{\ell\to\infty} X^\ell = \bar{X}$.

(ii) In the case that $\lim_{\ell\to\infty} \varepsilon_\ell = \varepsilon > 0$ we use the functional $\mathcal{J}_\varepsilon$ introduced in (6.8). Observe that $J_\varepsilon : \mathbb{R}^n \to \mathbb{R}_+$ is differentiable, convex, and *absolutely symmetric*, see Section 7.2. By Proposition 7.4, we therefore have, for X with singular value decomposition $X = U \operatorname{diag}(\sigma(X))V^*$,

$$\nabla \mathcal{J}_\varepsilon(X) = U \operatorname{diag}(j'_\varepsilon(\sigma_i(X)))V^*,$$

where the derivative of j_ε is given by

$$j_\varepsilon'(u) = \begin{cases} \operatorname{sgn}(u), & |u| \geq \varepsilon, \\ u/\varepsilon, & |u| < \varepsilon. \end{cases}$$

We have

$$\bar{X} = \bar{X}(\varepsilon) \in \arg\min_{\mathscr{S}(X)=\mathscr{M}} \mathcal{J}_\varepsilon(X)$$

if and only if $\langle \nabla \mathcal{J}_\varepsilon(\bar{X}), H \rangle = 0$ for all $H \in \ker \mathscr{S}$. It is now straightforward to verify that the latter is equivalent to

$$\langle W\bar{X}, H \rangle = 0, \text{ for all } H \in \ker \mathscr{S}, \tag{6.10}$$

where $W = [((\bar{X}\bar{X}^*)^{1/2})_\varepsilon]^{-1}$. Here, we recall that Z_ε is the ε-regularization of Z, see (2.3).

Since X^ℓ is a bounded sequence, it has accumulation points. Let us first show that any accumulation point of X^ℓ is a minimizer of $\mathcal{J}_\varepsilon$ subject to the constraint $\mathscr{S}(X) = \mathscr{M}$. Note that $\mathcal{J}_\varepsilon$ is not *strictly* convex, so such a minimizer is not necessarily unique. Let $(X^{\ell_j})_{j \in \mathbb{N}}$ be any convergent subsequence and $\tilde{X}$ its limit. Note that $W^{\ell_j} = [((X^{\ell_j}(X^{\ell_j})^*)^{1/2})_{\varepsilon_{\ell_j}}]^{-1}$. Hence, W^{ℓ_j} depends continuously on X^{ℓ_j} so that also W^{ℓ_j} converges to a limit $\tilde{W} = \lim_{j \to \infty} W^{\ell_j} = [((\tilde{X}\tilde{X}^*)^{1/2})_\varepsilon]^{-1}$. By invoking Proposition 6.1 (iii) we have also $X^{\ell_j+1} \to \tilde{X}$ for $j \to \infty$, hence by (5.4)

$$\langle \tilde{W}\tilde{X}, H \rangle = \lim_{j \to \infty} \langle W^{\ell_j} X^{\ell_j+1}, H \rangle = 0, \text{ for all } H \in \ker \mathscr{S}.$$

This is exactly the optimality condition (6.10) and therefore $\tilde{X} = \bar{X}$ is a minimizer of $\mathcal{J}_\varepsilon(X)$.

We now prove the error estimate. To begin, let X^{ℓ_j} be a converging subsequence of X^ℓ with limit $\tilde{X}$, which, for simplicity we denote again by X^ℓ. As just outlined $\tilde{X} = \bar{X}$ is a minimizer of $\mathcal{J}_\varepsilon$ subject to $\mathscr{S}(\tilde{X}) = \mathscr{M}$. Note that for any matrix X satisfying the constraint $\mathscr{S}(X) = \mathscr{M}$, and for any minimizer $\bar{X}$ of $\mathcal{J}_\varepsilon$ subject to the constraint $\mathscr{S}(\bar{X}) = \mathscr{M}$,

$$\|\bar{X}\|_* \leq \mathcal{J}_\varepsilon(\bar{X}) \leq \mathcal{J}_\varepsilon(X) \leq \|X\|_* + n\varepsilon,$$

where we have used the fact that $\bar{X}$ is a minimizer of $\mathcal{J}_\varepsilon$, so that

$$\|\bar{X}\|_* - \|X\|_* \leq n\varepsilon. \tag{6.11}$$

By the inverse triangle inequality, Lemma 6.6, and (6.11)

$$\|\bar{X} - X\|_* \leq \frac{1+\eta}{1-\eta}(\|\bar{X}\|_* - \|X\|_* + 2\rho_k(X)_*) \leq \frac{1+\eta}{1-\eta}(n\varepsilon + 2\rho_k(X)_*). \tag{6.12}$$

Thus we reach

$$n\varepsilon = \lim_{\ell \to \infty} n\varepsilon_\ell \leq \lim_{\ell \to \infty} \sigma_{K+1}(X^\ell) = \sigma_{K+1}(\bar{X}).$$

It follows from Proposition 7.2 that

$$(K + 1 - k)n\varepsilon \leq (K + 1 - k)\sigma_{K+1}(\bar{X}) \leq \|\bar{X} - X\|_* + \rho_k(X)_*$$

$$\leq \frac{1+\eta}{1-\eta}(n\varepsilon + 2\rho_k(X)_*) + \rho_k(X)_*.$$

Under the assumption that $K - k > \frac{2\eta}{1-\eta}$, this inequality yields

$$n\varepsilon \leq \frac{4(1+\eta)}{(1-\eta)((K-k)(1-\eta) - 2\eta)}\rho_k(X)_*.$$

Plugging this back into (6.12), we arrive at the desired result. $\square$

6.4. Discussion of related work. While we were finishing this paper, Maryam Fazel informed us of her joint work with Karthik Mohan, where a similar iteratively least squares minimization algorithm is studied [26]. We shortly outline the differences between our contribution and [26]. In [26] a direct and faithful generalization of the algorithm analyzed in [10] for sparse vector recovery is provided. In particular, instead of (2.12)

$$W^\ell = U^\ell \big(\operatorname{diag}(\max\{\sigma_j^\ell, \varepsilon_\ell\}) \big)^{-1} (U^\ell)^*,$$

the update rule for the weights is given there by

$$W^\ell = U^\ell \operatorname{diag}(((\sigma_j^\ell)^2 + \varepsilon_\ell^2)^{-1/2})(U^\ell)^*.$$

On the one hand, the drawback of our up-date rule (2.12) is that is it not anymore equivalent to an unconstrained minimization of the energy

$$\widetilde{\mathcal{J}}(X^\ell, W) = \mathcal{J}(X^\ell, W) + \frac{\varepsilon_\ell}{2}\|W^{1/2}\|_F^2,$$

with respect to the second variable $W \succ 0$, but to the constrained minimization of $\mathcal{J}(X^\ell, \cdot)$ as indicated in (5.3). Such equivalence had to be proven in Lemma 5.2 and Proposition 5.3, representing an additional technical difficulty in order to develop a proof of convergence, and introducing an element of novelty with respect to [10]. On the other hand, our choice is motivated by a significant improvement of the complexity of the algorithm when the operator $\mathscr{S}$ is *separable*, i.e., acts columnwise,

$$\mathscr{S}(X) = (\mathscr{S}_1 X_1, \ldots, \mathscr{S}_p X_p) = (\mathscr{M}_1, \ldots, \mathscr{M}_p) = \mathscr{M} \in \mathbb{C}^m,$$

where X_i, $i = 1, \ldots, p$, are the columns of X, and $\mathscr{S}_i$, $i = 1, \ldots, p$ are suitable matrices acting on the vectors X_i. We repeat that this case includes relevant situations such as the matrix completion problem. As clarified in Section 4.2.1, the Woodbury formula, which can applied to the matrices $\mathscr{S}_i(W^\ell)^{-1}\mathscr{S}_i^*$, thanks to the particular structure of the matrices W^ℓ as in (2.12), allows for their inversion with a much more economical cost than it would be possible by using the up-date rule proposed in [26]. Moreover, while the analysis in [26] also partially addresses reweighted least squares algorithms for Schatten q-norms, for $q < 1$, in the case of the nuclear norm ($q = 1$) the authors prove in [26, Theorem II.7] only an analogous version of our Theorem 6.11 (i), and do not develop a counterpart to (ii) and (iii).

7. Appendix.

7.1. Basic properties of singular values. In [37] Weyl proved the following stability estimate on the singular values.

THEOREM 7.1. *Let $X, Y \in M_{n \times p}$ be fixed. Then*

$$|\sigma_i(X) - \sigma_i(Y)| \leq \|X - Y\|_F, \quad i = 1, 2, \ldots.$$

PROPOSITION 7.2. *If $X \in M_{n \times p}$ and $Y \in M_{n \times p}$, then for any j, we have*

$$|\|X - X_{[j]}\|_* - \|Y - Y_{[j]}\|_*| \leq \|X - Y\|_*,$$

and for any $J > j$, we have $(J - j)\sigma_J(X) \leq \|X - Y\|_* + \|Y - Y_{[j]}\|_*$.

Proof. The first statement is a simple chain of inequalities:

$$\|X - X_{[j]}\|_* \leq \|X - Y_{[j]}\|_* \leq \|X - Y\|_* + \|Y - Y_{[j]}\|_*.$$

The result follows as we may reverse the roles of X and Y. For the second inequality, it suffices to note that $(J - j)\sigma_J(X) \leq \|X - X_{[j]}\|_*$. $\square$

7.2. Differentiation and singular values. Let $\mathcal{P}_n$ denote the set of all $n \times n$ permutation matrices. A function $f : \mathbb{R}^n \to \mathbb{R}$ is called absolutely symmetric if

$$f(PDx) = f(x) \quad \text{for all } x \in \mathbb{R}^n, P \in \mathcal{P}_n,$$

and for all diagonal matrices D having only the values $+1, -1$ on the diagonal. A function $F : M_{n \times p} \to \mathbb{R}^n$ is called unitarily invariant if $F(U^*XV)$ for all $X \in M_{n \times p}$, and for all unitary matrices $U \in M_{n \times n}, V \in M_{p \times p}$. The following fact is well-known, see for instance [24, Proposition 5.1].

PROPOSITION 7.3. *A function $F : M_{n \times p} \to \mathbb{R}$, $n \leq p$, is unitarily invariant if and only if $F(X) = f(\sigma(X))$ for some absolutely symmetric function $f : \mathbb{R}^n \to \mathbb{R}$. Here, $\sigma(X) \in \mathbb{R}^n$ denotes the vector of singular values of X.*

Let $f : \mathbb{R}^n \to \mathbb{R}$ be absolutely symmetric. Then the function $F(X) = f(\sigma(X))$ is convex if and only if f is convex. We have the following result concerning differentiation of F [24, Proposition 6.2].

PROPOSITION 7.4. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be an absolutely symmetric and convex function. The function $F = f \circ \sigma$ on $M_{n \times p}$ is differentiable at $X \in M_{n \times p}$ if and only if f is differentiable at $\sigma(X)$. In this case*

$$\nabla F(X) = U \operatorname{diag}(\nabla f(\sigma(X))) V^*,$$

where $X = U \operatorname{diag}(\sigma(X)) V^*$ for unitary matrices U, V.

REFERENCES

[1] N. Ailon and E. Liberty. An almost optimal unrestricted fast Johnson-Lindenstrauss transform. In D. Randall, editor, *SODA*, pages 185–191. SIAM, 2011.

[2] R. Bhatia. *Matrix Analysis*, volume 169 of *Graduate Texts in Mathematics*. Springer-Verlag, 1997.

[3] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge Univ. Press, 2004.

[4] J.-F. Cai, E. J. Candès, and Z. Shen. A singular value thresholding algorithm for matrix completion. *SIAM J. Optim.*, 20(4):1956–1982, 2010.

[5] E. Candès and Y. Plan. Tight oracle bounds for low-rank matrix recovery from a minimal number of random measurements. *IEEE Trans. Inform. Theory*, 57(4):2342–2359, 2011.

[6] E. J. Candès and B. Recht. Exact matrix completion via convex optimization. *Found. Comput. Math.*, 9:717–772, 2009.

[7] E. J. Candes and T. Tao. The power of convex relaxation: near-optimal matrix completion. *IEEE Trans. Inform. Theory*, 56(5):2053–2080, 2010.

[8] A. Chambolle and P.-L. Lions. Image recovery via total variation minimization and related problems. *Numer. Math.*, 76:167–188, 1997.

[9] A. Cohen, W. Dahmen, and R. A. DeVore. Compressed sensing and best k-term approximation. *J. Amer. Math. Soc.*, 22(1):211–231, 2009.

[10] I. Daubechies, R. A. DeVore, M. Fornasier, and C. S. Güntürk. Iteratively re-weighted least squares minimization for sparse recovery. *Comm. Pure Appl. Math.*, 63(1):1–38, 2010.

[11] M. Fazel. *Matrix Rank Minimization with Applications*. PhD thesis, Stanford University, 2002.

[12] M. Fornasier and H. Rauhut. Compressive sensing. In O. Scherzer, editor, *Handbook of Mathematical Methods in Imaging*, pages 187–228. Springer, 2011.

[13] S. Foucart. A note on guaranteed sparse recovery via ℓ_1-minimization. *Appl. Comput. Harmon. Anal.*, 29(1):97–103, 2010.

[14] S. Foucart and M. Lai. Sparsest solutions of underdetermined linear systems via ℓ_q-minimization for $0 < q \leq 1$. *Appl. Comput. Harmon. Anal.*, 26(3):395–407, 2009.

[15] D. Goldfarb and S. Ma. Convergence of fixed point continuation algorithms for matrix rank minimization. *Found. Comput. Math.*, 11(2):183–210, 2011.

[16] R. Gribonval and M. Nielsen. Sparse representations in unions of bases. *IEEE Trans. Inform. Theory*, 49(12):3320–3325, 2003.

[17] D. Gross. Recovering low-rank matrices from few coefficients in any basis. *preprint*, 2009.

[18] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert. Quantum state tomography via compressed sensing. *preprint*, 2009.

[19] R. Horn and C. Johnson. *Matrix Analysis*. Cambridge University Press, 1990.

[20] R. H. Keshavan, A. Montanari, and S. Oh. Matrix Completion from a Few Entries. *IEEE Trans. Inform. Theory*, 56:2980–2998, 2010.

[21] R. H. Keshavan, A. Montanari, and S. Oh. Matrix Completion from Noisy Entries. *J. Machine Learning Res.*, 11:2057–2078, 2010.

[22] F. Krahmer and R. Ward. New and improved Johnson-Lindenstrauss embeddings via the Restricted Isometry Property. *SIAM J. Math. Anal.*, 43(3):1269–1281, 2011.

[23] K. Lee and Y. Bresler. ADMiRA: Atomic decomposition for minimum rank approximation. *IEEE Trans. Inform. Theory*, 56(9):4402–4416, 2010.

[24] A. S. Lewis and H. S. Sendov. Nonsmooth analysis of singular values. I: Theory. *Set-Valued Analysis*, 13(3):213–241, 2005.

[25] Q. Mo and S. Li. New bounds on the restricted isometry constant δ_{2k}. *Appl. Comput. Harmon. Anal.*, in press, 2011. DOI: 10.1016/j.acha.2011.04.005.

[26] K. Mohan and M. Fazel. Iterative reweighted least squares for matrix rank minimization. In *Proceedings of the Allerton Conference*, pages 653–661, 2010.

[27] K. Mohan and M. Fazel. Reweighted nuclear norm minimization with application to system identification. In *Proceedings of the American Control Conference*, pages 2953–2959, 2010.

[28] B. K. Natarajan. Sparse approximate solutions to linear systems. *SIAM J. Comput.*, 24:227–234, 1995.

[29] S. Oymak, K. Mohan, M. Fazel, and B. Hassibi. A simplified approach to recovery conditions for low-rank matrices. *Preprint*, 2011.

[30] A. M. Pinkus. *On L_1-approximation*. Cambridge Tracts in Mathematics, 93. Cambridge: Cambridge University Press., 1989.

[31] A. M. Pinkus. Sparse representations and approximation theory. *J. Approx. Theory*, 163(3):388–412, 2011.

[32] H. Rauhut. Compressive sensing and structured random matrices. In M. Fornasier, editor, *Theoretical Foundations and Numerical Methods for Sparse Recovery*, Radon Series Comp. Appl. Math, pages 1–92. deGruyter, 2010.

[33] B. Recht. A simpler approach to matrix completion. *J. Machine Learning Res.*, to appear.

[34] B. Recht, M. Fazel, and P. Parillo. Guaranteed minimum rank solutions to linear matrix equations via nuclear norm minimization. *SIAM Rev.*, 52(3):471–501, 2010.

[35] B. Recht, B. Hassibi, and W. Xu. Null space conditions and thresholds for rank minimization. *Mathematical Programming. Ser B*, 127:175–211, 2011.

[36] R. Ward. Compressed sensing with cross validation. *IEEE Trans. Inform. Theory*, 55:5773–5782, 2009.

[37] H. Weyl. Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen (mit einer Anwendung auf die Theorie der Hohlraumstrahlung). *Math. Ann.*, 71:441–479, 1912.

[38] M. A. Woodbury. *Inverting modified matrices*. Statistical Research Group, Memo. Rep. no. 42. Princeton University, Princeton, N. J., 1950.